

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 8, 2005, 11:28:00 ; Search time 122.665 Seconds  
(without alignments)  
8481.439 Million cell updates/sec

Title: US-10-754-437-3

Perfect score: 22  
Sequence: 1 gattctgttcgctgcctcctg 22

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_hgt:\*  
3: gb\_in:\*  
4: gb\_cm:\*  
5: gb\_ov:\*  
6: gb\_pac:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sra:\*  
12: gb\_ey:\*  
13: gb\_un:\*  
14: gb\_vl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	22	100.0	296	10	S74899
2	22	100.0	363	11	BV089294
3	22	100.0	363	11	BV097602
4	22	100.0	463	10	DB3064
5	22	100.0	695	11	BV097603
6	22	100.0	695	11	BV097603
7	22	100.0	727	10	MUSMAMEB
8	22	100.0	746	4	AF215889
9	22	100.0	752	4	AF215889
10	22	100.0	757	4	BOVAM1GNII
11	22	100.0	798	10	BC059090
12	22	100.0	799	10	MUSMAMEA
13	22	100.0	820	4	BOVAM1GNII
14	22	100.0	5562	9	AB091787
15	22	100.0	6442	9	AB091783
16	22	100.0	6451	4	AB091789
17	22	100.0	8004	9	AB091786
18	22	100.0	9384	10	AF294397
19	22	100.0	95826	10	AL805974

20	20.4	92.7	65	6	CO531502
21	20.4	92.7	399	10	RU07054
22	20.4	92.7	457	10	RU060564
23	20.4	92.7	476	6	AR452534
24	20.4	92.7	500	6	AR077245
25	20.4	92.7	500	6	AR120190
26	20.4	92.7	500	6	BD107909
27	20.4	92.7	649	9	BC074951
28	20.4	92.7	659	9	BC069118
29	20.4	92.7	753	10	RU060562
30	20.4	92.7	765	4	AB032194
31	20.4	92.7	778	4	SSU43405
32	20.4	92.7	780	10	U01245
33	20.4	92.7	789	4	AB032193
34	20.4	92.7	793	6	AR077243
35	20.4	92.7	793	6	AR120188
36	20.4	92.7	793	6	CO731756
37	20.4	92.7	793	6	BD107907
38	20.4	92.7	793	9	HUMAMGXA
39	20.4	92.7	812	10	RU067130
40	20.4	92.7	825	10	RU051195
41	20.4	92.7	835	9	AF436849
42	20.4	92.7	885	10	CP0012200
43	20.4	92.7	5684	9	AB091785
44	20.4	92.7	5712	4	AB091793
45	20.4	92.7	6465	9	AB091781

#### ALIGNMENTS

RESULT 1  
LOCUS S74899 296 bp mRNA ROD 09-MAY-1995  
DEFINITION amelogenin {M194 transcript, alternatively spliced} [mice, enamel organ epithelia, mRNA Partial, 296 nt].  
ACCESSION S74899  
VERSION S74899.1 GI:802018  
KEYWORDS  
SOURCE Mus sp.  
ORGANISM Mus sp.  
REFERENCE 1 (bases 1 to 296)  
AUTHORS Simmer,J.P., Hu,C.C., Lau,B.C., Sarte,P., Slavkin,H.C. and Fincham,A.G.  
TITLE Alternative splicing of the mouse amelogenin primary RNA transcript  
JOURNAL Calcif. Tissue Int. 55 (4), 302-310 (1994)  
MEDLINE 95120737  
PubMed 7820782  
REMARK Genbank staff at the National Library of Medicine created this entry [NCBI glibseq 159650] from the original journal article.  
FEATURES  
source 1..296  
/organism="Mus sp."  
/mol\_type="mRNA"  
/db\_xref="taxon:10095"  
1..296  
/gene="amelogenin"  
69..296  
/gene="amelogenin"  
/codon\_start=1  
/product="amelogenin"  
/protein\_id="AAB33093.1"  
/db\_xref="GI:802018"  
/translation="MGTWILFACLLGAAPAMPUPPHGSPGYINLSYKSHSQAINTD  
RTALVITPKWQSMIRQPPSYGVRPMGWL"

#### ORIGIN

Query Match 100.0% ; Score 22; DB 10; Length 296;  
Best Local Similarity 100.0% ; Pred. No. 1.9;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Oy      1 GATTTGTTGCTGCTCCTG 22
        |||
        80 GATTTGTTGCTGCTCCTG 101

RESULT 2
BV089294/c      363 bp      DNA      linear      STS 15-OCT-2003
LOCUS      RPAMMSE0001190 Roche Palo Alto Mus musculus STS genomic, sequence
DEFINITION      tagged site.
ACCESSION      BV089294
VERSION      BV089294.1 GI:37666773
KEYWORDS
SOURCE      Mus musculus (house mouse)
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Scurionath; Muridae; Murinae; Mus.
REFERENCE      1 (bases 1 to 363)
AUTHORS      Usuka,J., Liao,G., Cheng,J., Nguyen,A., Bach,C., Puech,A.,
McPherson,J.D., Foernzler,D. and Peltz,G.
TITLE      Mus musculus SNPs
JOURNAL      Unpublished (2003)

COMMENT
Contact: Jonathan Usuka
Roche Palo Alto Genetics and Genomics Department
Roche Palo Alto
3431 Hillview Ave, Mailstop S3-1, Palo Alto, CA 94024, USA
Tel: 6508555807
Email: jonathan.usuka@roche.com
Primer A: No primer submitted
Primer B: No primer submitted.
Location/Qualifiers
1..363
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/map="X-73268-72907-AL005974.7.1.95826"
/clone_lib="Roche Palo Alto"
/notes="SNPs developed from assay sequences derived from 15
different strains of mice (as of October 1, 2003). Those
strains include A/J, A/HeJ, 129/Sv, AKR/J, B10.D2-H2/GSDJ,
BALB/cByJ, BALB/cJ, C3H/HeJ, C57BL/6J, CAST/Ei, DBA/2J,
MRL/MpJ, NZB/BinJ, NZW/Lac, SPRET/Ei."
<1..>363

ORIGIN
STS
Query Match      100.0%; Score 22; DB 11; Length 363;
Best Local Similarity 100.0%; Pred. No. 1.9;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      1 GATTTGTTGCTGCTCCTG 22
        |||
        134 GATTTGTTGCTGCTCCTG 113

RESULT 3
BV097602/c      363 bp      DNA      linear      STS 14-JAN-2004
LOCUS      RPAMMSE00010112 Roche Palo Alto Mus musculus STS genomic, sequence
DEFINITION      tagged site.
ACCESSION      BV097602
VERSION      BV097602.1 GI:40810678
KEYWORDS
SOURCE      Mus musculus (house mouse)
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Scurionath; Muridae; Murinae; Mus.
REFERENCE      1 (bases 1 to 363)
AUTHORS      Usuka,J., Liao,G., Cheng,J., Nguyen,A., Bach,C., Puech,A.,
McPherson,J.D., Foernzler,D. and Peltz,G.
TITLE      Mus musculus SNPs
JOURNAL      Unpublished (2003)

COMMENT
Contact: Jonathan Usuka
Roche Palo Alto Genetics and Genomics Department
Roche Palo Alto
3431 Hillview Ave, Mailstop S3-1, Palo Alto, CA 94024, USA
Tel: 6508555807
Email: jonathan.usuka@roche.com
Primer A: No primer submitted with this STS
Primer B: No primer submitted with this STS.
Location/Qualifiers
1..363
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/map="X-73268-72907-AL005974.7.1.95826"
/clone_lib="Roche Palo Alto"
/notes="SNPs developed from assay sequences derived from 15
different strains of mice (as of October 1, 2003). Those
strains include A/J, A/HeJ, 129/Sv, AKR/J, B10.D2-H2/GSDJ,
BALB/cByJ, BALB/cJ, C3H/HeJ, C57BL/6J, CAST/Ei, DBA/2J,
MRL/MpJ, NZB/BinJ, NZW/Lac, SPRET/Ei."
<1..>363

ORIGIN
STS
Query Match      100.0%; Score 22; DB 11; Length 363;
Best Local Similarity 100.0%; Pred. No. 1.9;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      1 GATTTGTTGCTGCTCCTG 22
        |||
        134 GATTTGTTGCTGCTCCTG 113

RESULT 4
D83063S2      463 bp      DNA      linear      ROD 26-NOV-2003
LOCUS      Mus musculus DNA for amelogenin, exon 2.
DEFINITION      D83064
ACCESSION      D83064.1 GI:2687869
VERSION
KEYWORDS
SEGMENT      2 of 5
SOURCE      Mus musculus (house mouse)
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Scurionath; Muridae; Murinae; Mus.
REFERENCE      1
AUTHORS      Oida,S., Miyazaki,H., Iimura,T., Suzuki,M., Sasaki,S. and
Shimokawa,H.
TITLE      Molecular structure of the mouse amelogenin genomic DNA
JOURNAL      DNA Seq. 6 (5), 307-310 (1996)
MEDLINE      97142134
PUBMED      8988368
REFERENCE      2 (bases 1 to 463)
AUTHORS      Oida,S.
TITLE      Direct Submission
JOURNAL      Submitted (14-JAN-1996) Shinichiro Oida, Tokyo Medical and Dental
University, Fac. Dentistry, Dept. Biochemistry; 1-5-45 Yushima,
Bunkyo-ku, Tokyo 113, Japan (E-mail:s.oida.obh@dent.tmd.ac.jp,
Tel:03-5803-5448, Fax:03-5803-0187)
On Dec 15, 1997 this sequence version replaced gi:2662357.
Location/Qualifiers
1..463
/organism="Mus musculus"
/mol_type="genomic DNA"
/sub_species="Balb-c"
/db_xref="taxon:10090"
/chromosome="X"
/map="XF3-F4"
/clone="NAMEG-2"
/cell_type="leukocyte"
/clone_lib="EMBL-3 SP6-T7"
/dev_stage="adult"
Join(D83063.1:203..258,222..233)
<1..221
5'UTR
Intron

```



exon /number=1  
222 .287  
/product="amelogenin"

Intron /number=2  
288 .463  
/number=2

## ORIGIN

Query Match 100.0%; Score 22; DB 10; Length 463;  
Best Local Similarity 100.0%; Pred. No. 2;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATTTGTTGCTGCTCCTG 22  
|||||  
Db 245 GATTTGTTGCTGCTCCTG 266

RESULT 5  
BV089295/c 695 bp DNA linear STS 15-OCT-2003  
LOCUS RPAMMSE0001191 Roche Palo Alto Mus musculus STS genomic, sequence  
DEFINITION tagged site.

ACCESSION BV089295  
VERSION BV089295  
KEYWORDS GI:3766774  
SOURCE Mus musculus (house mouse)

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 695)  
Uauka,J., Liao,G., Cheng,J., Nguyen,A., Bach,C., Puech,A.,  
McPherson,J.D., Foernzler,D. and Peltz,G.

REFERENCE Mus musculus SNPs  
JOURNAL Unpublished (2003)

## TITLE

COMMENT

Contact: Jonathan Uauka  
Roche Palo Alto Genetics and Genomics Department  
Roche Palo Alto  
3431 Hillview Ave, Mailstop S3-1, Palo Alto, CA 94024, USA  
Tel: 6508555807  
Email: Jonathan.Uauka@roche.com  
Primer A: No primer submitted  
Primer B: No primer submitted.

## FEATURES

source

1..695  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/map="X-73613-72919-AL805974.7.1.95826"  
/clone\_11b="Roche Palo Alto"  
/note="SNPs developed from assay sequences derived from 15  
different strains of mice (as of October 1, 2003). Those  
strains include A/J, A/HeJ, -129/Sv, AKR/J, B10.D-H2/OSuJ,  
BALB/cByJ, BALB/cJ, C3H/HeJ, C57BL/6J, CAST/Ei, DBA/2J,  
MRL/MpJ, NZB/BinJ, NZM/Lac, SPRET/Ei.."  
<1..>695

## ORIGIN

STS

Query Match 100.0%; Score 22; DB 11; Length 695;  
Best Local Similarity 100.0%; Pred. No. 2;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATTTGTTGCTGCTCCTG 22  
|||||  
Db 479 GATTTGTTGCTGCTCCTG 458

RESULT 6  
BV097603/c 695 bp DNA linear STS 14-JAN-2004  
LOCUS RPAMMSE00010113 Roche Palo Alto Mus musculus STS genomic, sequence  
DEFINITION tagged site.  
ACCESSION BV097603

VERSION BV097603.1 GI:40810679  
KEYWORDS STS.  
SOURCE Mus musculus (house mouse)

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 695)  
Uauka,J., Liao,G., Cheng,J., Nguyen,A., Bach,C., Puech,A.,  
McPherson,J.D., Foernzler,D. and Peltz,G.

REFERENCE Mus musculus SNPs  
JOURNAL Unpublished (2003)

## TITLE

COMMENT

Contact: Jonathan Uauka  
Roche Palo Alto Genetics and Genomics Department  
Roche Palo Alto  
3431 Hillview Ave, Mailstop S3-1, Palo Alto, CA 94024, USA  
Tel: 6508555807  
Email: Jonathan.Uauka@roche.com  
Primer A: No primer submitted with this STS  
Primer B: No primer submitted with this STS.

## FEATURES

source

1..695  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/map="X-73613-72919-AL805974.7.1.95826"  
/clone\_11b="Roche Palo Alto"  
/note="SNPs developed from assay sequences derived from 15  
different strains of mice (as of October 1, 2003). Those  
strains include A/J, A/HeJ, 129/Sv, AKR/J, B10.D-H2/OSuJ,  
BALB/cByJ, BALB/cJ, C3H/HeJ, C57BL/6J, CAST/Ei, DBA/2J,  
MRL/MpJ, NZB/BinJ, NZM/Lac, SPRET/Ei.."  
<1..>695

## ORIGIN

STS

Query Match 100.0%; Score 22; DB 11; Length 695;  
Best Local Similarity 100.0%; Pred. No. 2;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATTTGTTGCTGCTCCTG 22  
|||||  
Db 479 GATTTGTTGCTGCTCCTG 458

RESULT 7  
MUSMAMB 727 bp mRNA linear ROD 04-FEB-2003  
LOCUS Mus musculus MAMB mRNA for amelogenin, complete cds.  
DEFINITION D31769  
VERSION D31769.1 GI:1321652

KEYWORDS Mus musculus (house mouse)

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1  
Snead,M.L., Lau,E.C., Zeichner-David,M., Fincham,A.G., Woo,S.L. and  
Slavkin,H.C.

TITLE DNA sequence for cloned cDNA for murine amelogenin reveal the amino  
acid sequence for amel-specific protein  
JOURNAL Biochem. Biophys. Res. Commun. 129 (3), 812-818 (1985)

MEDLINE 85251692  
PUBMED 4015654

## REFERENCE

AUTHORS

## TITLE

Alternative splicing of the mouse amelogenin primary RNA transcript  
contributes to amelogenin heterogeneity  
JOURNAL Biochem. Biophys. Res. Commun. 188 (3), 1253-1260 (1992)  
MEDLINE 93075222  
PUBMED 1445358  
REFERENCE 3 (bases 1 to 727)



REFERENCE	AUTHORS	TITLE
1	(bases 1 to 752)	Caprine; Capra
2	(bases 1 to 752)	Chen, C.-M., Cheng, W.T.K., Chang, C.-C., Weng, T.-L., Jiang, Y.-N. and Heu, J.-T.
3	(bases 1 to 752)	Cloning, cDNA sequence, and alternative splicing of goat amelogenin mRNAs
4	(bases 1 to 752)	Unpublished
5	(bases 1 to 752)	Chen, C.-M., Cheng, W.T.K., Chang, C.-C., Weng, T.-L., Jiang, Y.-N. and Heu, J.-T.
6	(bases 1 to 752)	Direct Submission
7	(bases 1 to 752)	Submitted (13-DEC-1999) Department of Zoology, Life Science College, No. 250, Kuo Kuang Road, Tainchung 402, Taiwan
8	(bases 1 to 752)	Location/Qualifiers
9	(bases 1 to 752)	1. .752
10	(bases 1 to 752)	/organism="Capra hircus"
11	(bases 1 to 752)	/mol_type="mRNA"
12	(bases 1 to 752)	/db_xref="taxon:9925"
13	(bases 1 to 752)	/chromosome="X"
14	(bases 1 to 752)	/clone="AML-X-10"
15	(bases 1 to 752)	/cell_type="ameloblast"
16	(bases 1 to 752)	/note="bred: Alpine"
17	(bases 1 to 752)	<1. .752
18	(bases 1 to 752)	/gene="AML-X"
19	(bases 1 to 752)	<1. .34
20	(bases 1 to 752)	/gene="AML-X"
21	(bases 1 to 752)	35. .658
22	(bases 1 to 752)	/gene="AML-X"
23	(bases 1 to 752)	/codon_start=1
24	(bases 1 to 752)	/product="amelogenin precursor"
25	(bases 1 to 752)	/protein_id="AAG4396.1"
26	(bases 1 to 752)	/db_xref="GI:12004312"
27	(bases 1 to 752)	/translation="MGTWILFACILGAASFMPRLPHGHPGYNFSYEVPTLKYOSI
28	(bases 1 to 752)	MIRBPSYGYEYPMGMLHQLIIVNSQOTPNALDPHHIIIPVPAQDPVPOQPMW
29	(bases 1 to 752)	PVPCOHMTPTGCHQNPRLPLPAQCPQOSIQPHNQIDQLDQMLQPLQPLQPLQ
30	(bases 1 to 752)	PVHTQPLPQPLPLPLPLPMLPDLVLEDPATDKTKREVP"
31	(bases 1 to 752)	35. .82
32	(bases 1 to 752)	/gene="AML-X"
33	(bases 1 to 752)	/note="secretory signal peptide"
34	(bases 1 to 752)	659. .5752
35	(bases 1 to 752)	/gene="AML-X"
36	(bases 1 to 752)	/note="highly divergent from AML-Y 3' UTR"
37	(bases 1 to 752)	
38	(bases 1 to 752)	
39	(bases 1 to 752)	
40	(bases 1 to 752)	
41	(bases 1 to 752)	
42	(bases 1 to 752)	
43	(bases 1 to 752)	
44	(bases 1 to 752)	
45	(bases 1 to 752)	
46	(bases 1 to 752)	
47	(bases 1 to 752)	
48	(bases 1 to 752)	
49	(bases 1 to 752)	
50	(bases 1 to 752)	
51	(bases 1 to 752)	
52	(bases 1 to 752)	
53	(bases 1 to 752)	
54	(bases 1 to 752)	
55	(bases 1 to 752)	
56	(bases 1 to 752)	
57	(bases 1 to 752)	
58	(bases 1 to 752)	
59	(bases 1 to 752)	
60	(bases 1 to 752)	
61	(bases 1 to 752)	
62	(bases 1 to 752)	
63	(bases 1 to 752)	
64	(bases 1 to 752)	
65	(bases 1 to 752)	
66	(bases 1 to 752)	
67	(bases 1 to 752)	
68	(bases 1 to 752)	
69	(bases 1 to 752)	
70	(bases 1 to 752)	
71	(bases 1 to 752)	
72	(bases 1 to 752)	
73	(bases 1 to 752)	
74	(bases 1 to 752)	
75	(bases 1 to 752)	
76	(bases 1 to 752)	
77	(bases 1 to 752)	
78	(bases 1 to 752)	
79	(bases 1 to 752)	
80	(bases 1 to 752)	
81	(bases 1 to 752)	
82	(bases 1 to 752)	
83	(bases 1 to 752)	
84	(bases 1 to 752)	
85	(bases 1 to 752)	
86	(bases 1 to 752)	
87	(bases 1 to 752)	
88	(bases 1 to 752)	
89	(bases 1 to 752)	
90	(bases 1 to 752)	
91	(bases 1 to 752)	
92	(bases 1 to 752)	
93	(bases 1 to 752)	
94	(bases 1 to 752)	
95	(bases 1 to 752)	
96	(bases 1 to 752)	
97	(bases 1 to 752)	
98	(bases 1 to 752)	
99	(bases 1 to 752)	
100	(bases 1 to 752)	
101	(bases 1 to 752)	
102	(bases 1 to 752)	
103	(bases 1 to 752)	
104	(bases 1 to 752)	
105	(bases 1 to 752)	
106	(bases 1 to	

FEATURES	source	Location/Qualifiers
1..757		/organism="Bos taurus"
		/mol_type="mRNA"
		/db_xref="taxon:9913"
		/cell_type="ameloblast"
		/tissue_type="tooth"
		/dev_stage="fetus"
1..757		/gene="amelogenin"
41..619		/gene="amelogenin"
		/note="class II"
		/codon_start=1
		/product="amelogenin"
		/protein_id="AA30373.1"
		/db_xref="GI:162662"
		/translation="MGTVITIFACLLGAARSRMPRPHGHGICINFSYEVUTPLKRYOS MIRHPSPSYGIEPMGWLHIIQIIPVVSQTPONHALQPHHIIPIVPAQGVPPQPMW PVGQSHMPTQRIHQNPILPDAQOSFPQPIQPHLPQPPVPHIQRLLPPQPLP PIPPQPLPVLPLPDLPLEAMPADYDKTKEBEVD"
Query Match	100.0%	Score 22; DB 4; Length 757;
Best Local Similarity	100.0%	Pred. No. 2;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
Qy	1	GATTGTGTCGCGCTCCG 22
Db	52	GATTGTGTCGCGCTCCG 73
RESULT 11		
LOCUS BC059090		
DEFINITION Mus musculus cDNA clone MGC:70128 IMAGE:30286944, complete cds.		
ACCESSION BC059090		
VERSION BC059090.1		GI:37474841
KEYWORDS MGC.		
SOURCE Mus musculus (house mouse)		
ORGANISM Mus musculus		
REFERENCE Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, P.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Bluetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Heide, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Spatleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheer, T.E., Brownstein, M.J., Ustin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Kaha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwen, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Wierley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Huylk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Pahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shvachenko, Y., Bouckson, M.C., Rodriguez, A.C., Grimwood, J., Schultz, D., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalish, D.E., Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.		
TITLE Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences		
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)		
MEDLINE 23388257		
PUBMED 12477932		
REFERENCE 2 (bases 1 to 798)		
AUTHORS Strausberg, R.		
TITLE Direct Submission		

**JOURNAL**

**JOURNAL** Submitted (01-OCT-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

**REMARK** NIH-MGC Project URL: <http://mgc.nci.nih.gov>

**COMMENT** Contact: MGC help desk

REMARK  
COMMENT

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/BLINL at: <http://Image.llnl.gov>  
Series: IRRAK Plate: 131 Row: d Column: 8  
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer Frequency ORF analysis, Genomescan gene prediction.

```

source
1. .798
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="MGC:70128 IMAGE:30286944"
/tissue_type="jaw and limb, mouse, day 18.5 to new born,
pool of mature fore and hind limb, maxilla and mandible
containing endochondral and membranous bone, formed
joints, tendon, ligaments, dermis, epidermis, muscle and
teeth with newly forming dentin and enamel"
/clone_1ib="NIH_MGC_136"
/lab_host="DH10B"
/notes="Vector: pCMV-SPORT6.1"
39. .698

```

**misc\_feature**  
87..392  
/note="Amelogenin, Region: pfam02948, Amelogenin, Amelogenin. Amelogenins play a role in biomineralization. They seem to regulate the formation of crystallites during the secretory stage of tooth enamel development. Thought to play a major role in the structural organization and mineralization of developing enamel. They are found in the extracellular matrix. Mutations in X-chromosomal amelogenin can cause Amelogenesis imperfecta"  
/db\_xref="CDD:pfam02948"

Query Match	100.0%	Score 22	DB 10	Length 798
Best Local Similarity	100.0%	Prd. No. 2		
Matches	22	Conservative	0	Mismatches 0; Indels 0; Gaps 0;
QY	1	GATTTCGTTTGCCCTGCTCTG	22	
Db	50	GATTTCGTTTGCCCTGCTCTG	71	

## RESULT 12

MUSMAMEA LOCUS 799 bp mRNA linear ROD 04-FEB-2003  
 MUSMAMEA DEFINITION Mus musculus MAME mRNA for amelogenin, complete cds.  
 D11768  
 D31768.1 GI:1321650  
 \* Mus musculus (house mouse)  
 SOURCE  
 ORGANISM

REFERENCE	AUTHORS	TITLE	JOURNAL	MEDLINE	PUBMED	REFERENCE
1	Snead, M. L., Lau, E. C., Zeichner-David, M., Fincham, A. G., Woo, S. L. and Slavkin, H. C.	DNA sequence for cloned cDNA for murine amelogenin reveal the amino acid sequence for enamel-specific protein	Biochem. Biophys. Res. Commun.	129 (3), 812-818 (1985)	85251692	
					4015654	
2	Lau, E. C., Simmer, J. P., Bringas, P. Jr., Hsu, D. D., Hu, C. C.,					

COMMENT	On May 16, 1996 this sequence version replaced gi:577326.
FEATURES	
Location/Qualifiers	
1	
200	
Source	

```

exon
    /gene="NAME"
    1..47
    /gene="NAME"
    /number=1
    48..113
    /gene="NAME"
    /number=2
    60..650
    /gene="NAME"
    /note="enamel-specific protein"
    /codon_start=1
    /product="amelogenin precursor"
    /protein_id="AA06546.1"
    /db_xref="GI:1321651"
    /translation="MGTWILFACILGAFAFMPLEDPHFGSPGYINLSYEVLTPLMKYS
MIRQPVSYSGEPMGMLHQIILVLSQHPSPSTLQPHNHLPVPAQOQVPAPOQPM
VPRSHMSTPTQHQPNIPPSAQPFQOQFQOPQIPQSHQPMQPSPLHMQGLAQ
PLPLPLFSMQPLSPILBELPLFAWPAATDKTRBEVD"
    60..110
    /gene="NAME"
    111..647
    /gene="NAME"

```

sig_peptide	60..110
	/gene="MAME"
mat_peptide	111..647
	/gene="MAME"

```

exon
    /product="amelogenin"
    /note="enamel-specific protein"
    114. . 161
    /gene="WAME"
    /number=3
exon
    162. . 206
    /gene="WAME"
    /number=5
exon
    207. . 644
    /gene="WAME"
    /number=6
exon
    645. . 799
    /gene="WAME"
    /number=7
polyA_signal
    780. . 785
    /gene="WAME"
polyA_site
    799
    /gene="WAME"
ORIGIN
```

```

Query Match          100.0%; Score 22; DB 10; Length 799;
Beet Local Similarity 100.0%; Pred. No. 2;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GATTTCGTCGCCGCCCTCG 22
        |||||
Db       71 GATTTCGTCGCCCTCG 92

```

LOCUS	820 bp	mRNA	linear	MM 26-APR-1993
BOVAMLGN1				
DEFINITION	Bovine (class I) amelogenin mRNA, complete cds.			
ACCESSION	M63499 J05307			
VERSION	M63499.1 GI:162659			
KEYWORDS	amelogenin.			
SOURCE	Bos taurus (cow)			
ORGANISM	Bos taurus			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.			
AUTHORS	I (bases 1 to 820)			
	Gibson,C., Golub,E., Herold,R., Risser,M., Ding,W., Shimokawa,H., Young,M., Termino,J. and Rosenblom,J.			
TITLE	Structure and expression of the bovine amelogenin gene			
JOURNAL	Biochemistry 30 (4), 1075-1079 (1991)			
PMID	9113686			
COMMENT	1989679			
	Original source text: Bovine fetal teeth ameloblast (class I), cDNA to mRNA.			

FEATURES	SOURCE	Location/Qualifiers
gene	CDS	1..820
		location/Qualifiers
		1..820
		/organism="Bos taurus"
		/mol_type="mRNA"
		/db_xref="taxon:9913"
		/cell_type="ameloblast"
		/tissue_type="tooth"
		/dev_stage="fetus"
		1..820
		/gene="amelogenin"
		41..682
		/gene="amelogenin"
		/note="class I"
		/codon_start=1
		/product="amelogenin"
		/protein_id="AAA30372.1"
		/db_xref="GI:162660"
		/translation="MGTWILPACLLGAASMPDLPHPGHPGYNFSYVLTPLKMYQSS MIRHDPYSGYGEPMGSGMLHQLIPVSSQSTPQNALQDHNHILPMVPAQDPVPRQPMV FVSGHSMTPTQHNQNPILPLPAQDPFGQSGQDPQNPQLQDNPQNPQNPQLPLC PLQDPNPNPQNPQNPPLPPIPFMPQPLPMLPDLPLQAWPATDKTKEEVD"

Query Match	100.0%;	Score 22;	DB 4;	Length 820;
Best Local Similarity	100.0%;	Pred. No. 2;		
Matches	22;	Conservative 0;	Mismatches 0;	Indels 0;
				Gaps 0;
Qy	1	GATTGTTGTCCTGCTGCTG	22	
Db	52	GATTGTTGTCCTGCTGCTG	73	

RESULT 14		
LOCUS	AB091787	
DEFINITION	5562 bp DNA linear PRI 02-MAY-2003	
ACCESSION	AB091787	
VERSION	AB091787.1	GI:29126027
Oleleum garnetii AMELX gene for amelogenin, partial cds.		

SOURCE ORGANISM	Otolemur garnettii (small-eared galago)	Otolemur garnettii

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE

1 Iwase, M., Satta, Y., Hirai, Y., Hirai, H., Inai, H. and Takahata, N.  
From the Cover: The amelogenin loci span an ancient pseudautosomal  
boundary in diverse mammalian species  
Proc. Natl. Acad. Sci. U.S.A. 100 (9), 5258-5263 (2003)  
22608569

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
2 (bases 1 to 5562)  
Iwase, M., Satta, Y. and Takahata, N.  
Direct Submission  
Submitted (19-SEP-2002) Mineyo Iwase, Graduate University for  
Advanced Studies (Sokenda), Department of Bioystems Science;  
Shonan Kokusaiinmura, Hayama, Kanagawa 240-0193, Japan  
(E-mail: iwaseemin@koryu.w01.soken.ac.jp, Tel: 81-468-58-1571,  
Fax: 81-468-58-1544)

FEATURES	
source	Location/Qualifiers
gene	1..5562 /organism="Olelemur garnettii" /mol_type="genomic DNA" /db_xref="taxon:30611" /sex="male" 267..5562 /gene="AMELX" join(267..321,1401..1412) /gene="AMELX" join(1413..1466,3385..3432,4699..4740,4832..4876, 5125..>5562) /gene="AMELX" /codon_start=1 /product="amelogenin" /protein_id="BAC6107.1" /db_xref="GI:29126028" /translation="MGTWTFACILGAFAMRLPRPHGHGYINSEKYSQAINID KTLAVLTRYKYSQMLRPYRYSYGVEMGMLINQITPLYSQDHPRTNTIQRNNITRV VMAQQRVPOQRPMVRFQGSQSMRTIQDHNQSLRRPAQDRPQVQDQRPQDQRPQ MQRQIQRPQRVNRQRLRPLRLFGMRQLRPLRDLRLPLAMPAT"
5' UTR	
CDS	

```

Query Match      100.0 %; Score 22; DB 9; Length 5562;
Best Local Similarity 100.0%; Pred. No. 2.3;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GATTTGTGGCTGCTCCTCG 22
          |||||
Db      1424 GATTGTGGCTGCTCCTCG 1445

```

RESULT 15	AB091783	LOCUS	AB091783	6442 bp	DNA	linear	PRI 02-MAY-2003
DEFINITION	AB091783	Salmonella enterica serovar AMELX gene for amelogenin, partial cds.					
ACCESSION	AB091783						
VERSION	AB091783.1						GI:29126019

## KEYWORDS

SOURCE  
ORGANISM

Saimiri sciureus (common squirrel monkey)  
Saimiri sciureus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Cebinae;  
Saimiri.

## REFERENCE

AUTHORS  
TITLE

1 Iwase, M., Satra, Y., Hirai, Y., Hirai, H., Imai, H. and Takahata, N.  
From the cover: The amelogenin loci span an ancient pseudautosomal  
boundary in diverse mammalian species  
Proc. Natl. Acad. Sci. U.S.A. 100 (9), 5258-5263 (2003)

JOURNAL  
MEDLINE  
PUBMED

22608569  
12672962

## REFERENCE

AUTHORS  
TITLE  
JOURNAL

2 (bases 1 to 6442)  
Iwase, M., Satra, Y. and Takahata, N.  
Direct Submission  
Submitted (19-SEP-2002) Mineyo Iwase, Graduate University for  
Advanced Studies (Soken), Department of Biosystems Science,  
Shonan Kikusaijima, Hayama, Kanagawa 240-0193, Japan  
(E-mail: iwase@koryu.w01.soken.ac.jp, Tel: 81-468-58-1571,  
Fax: 81-468-58-1544)

## FEATURES

source

Location/Qualifiers

1..6442  
/organism="Saimiri sciureus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9521"  
/sex="male"  
936..6442  
/gene="AMELX"  
join(936..991,2298..2309)  
/gene="AMELX"  
join(2310..2363,4289..4336,5592..5633,5725..5769,  
6041..6442)  
/gene="AMELX"  
/codon\_start=1  
/product="amelogenin"  
/protein\_id="BAC6103.1"  
/db\_xref="GI:29126020"  
/translation="MGWTILFACLGAAFMPLPRPHGHPGYINSEYNSHQAINID  
RTGLVITPLKWTQSMIRPAVPSGYEPMGMLHHQITIVLSQHPPTHTLQPHHHIV  
VPAQOPVVPQOPMPPVPGQHSMTPTQHQPNIPLPPAQOPFQPTVQPPHPQPPQPP  
VHPMOPLPPOPPLPMPFPMQPLPMLPDLPLBWPAT"

gene

5'UTR

CDS

## ORIGIN

Query Match 100.0%; Score 22; DB 9; Length 6442;

Best Local Similarity 100.0%; Pred. No. 2.3; Mismatches 0; Indels 0; Gaps 0;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GATTGTGCTGCTGCTGCTG 22  
|||||  
Db 2321 GATTGTGCTGCTGCTGCTG 2342

Search completed: February 8, 2005, 16:36:40  
Job time : 126.665 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 8, 2005, 11:25:10 ; Search time 29.1839 Seconds  
(without alignments)  
3957.220 Million cell updates/sec

Title: US-10-754-437-3

Perfect score: 22  
Sequence: 1 gattctgttcgtcgtccctcg 22

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 413486 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_23sep04:\*  
1: geneseqn1980a:\*  
2: geneseqn1990a:\*  
3: geneseqn2000a:\*  
4: geneseqn2001a:\*  
5: geneseqn2001b:\*  
6: geneseqn2002a:\*  
7: geneseqn2002b:\*  
8: geneseqn2003a:\*  
9: geneseqn2003b:\*  
10: geneseqn2003c:\*  
11: geneseqn2003d:\*  
12: geneseqn2004a:\*

Prod. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	22	100.0	722	6	AAL41111
2	22	100.0	722	6	AAL41110
3	20.4	92.7	65	6	ABN28389
4	20.4	92.7	270	12	ADM80835
5	20.4	92.7	318	12	ADM80836
6	20.4	92.7	476	3	AAZ50832
7	20.4	92.7	549	12	ADM80826
8	20.4	92.7	556	10	ADB59026
9	20.4	92.7	556	10	ADB53782
10	20.4	92.7	623	12	ADM80827
11	20.4	92.7	750	2	AAZ07020
12	20.4	92.7	793	2	AAZ07018
13	20.4	92.7	259202	12	ADQ18492
14	20	90.9	20	2	AAK18413
15	19.4	88.2	2190	8	ABK16470
16	19.4	88.2	2190	11	ADK39263
17	19.4	88.2	2685	6	ABL67223
18	19.4	88.2	2685	10	ADB75542
19	19.4	88.2	4055	10	ADP81652
20	19.4	88.2	4181	12	ADL12637
21	19.4	88.2	4181	12	ADL12637

C	22	19	86.4	19	2	AAK18412	AAK18412 PCR prime
C	23	18.8	85.5	802	2	AAZ07019	AAZ07019 Amelogeni
C	24	18.8	85.5	852	12	ADQ22977	ADQ22977 Human BOF
C	25	18.4	83.6	1287	4	ABK41746	ABK41746 cDNA enco
C	26	18.4	83.6	1287	9	ADB59413	ADB59413 Connectiv
C	27	18.4	83.6	1295	4	ABK42097	ABK42097 cDNA enco
C	28	18.4	83.6	1295	9	ADB59764	ADB59764 Connectiv
C	29	17.8	80.9	173	3	AAK31387	AAK31387 Human sec
C	30	17.4	79.1	80	12	ADM95374	ADM95374 Rat anti
C	31	17.4	79.1	116	2	AAZ00965	AAZ00965 Pgi codin
C	32	17.4	79.1	3923	10	AAZ042229	AAZ042229 Toxicity
C	33	17.4	79.1	5324	2	AAZ00879	AAZ00879 Pgi splic
C	34	17.4	79.1	56516	2	AAZ00879	AAZ00879 Pgi genom
C	35	17.4	79.1	56520	2	AAZ01022	AAZ01022 Wild type
C	36	17.2	78.2	286	2	AAV89814	AAV89814 EST clone
C	37	17.2	78.2	493	6	ABL80978	ABL80978 Human ova
C	38	17.2	78.2	563	4	AAI16914	AAI16914 Probe #68
C	39	17.2	78.2	563	4	ABA60895	ABA60895 Human foe
C	40	17.2	78.2	563	4	AAI40790	AAI40790 Probe #94
C	41	17.2	78.2	563	4	ABA28878	ABA28878 Probe #73
C	42	17.2	78.2	563	4	AAK35074	AAK35074 Human bon
C	43	17.2	78.2	563	4	AAK09185	AAK09185 Human bra
C	44	17.2	78.2	563	4	ABS34826	ABS34826 Human liv
C	45	17.2	78.2	563	6	ABS09542	ABS09542 Human gen

## ALIGNMENTS

RESULT 1	AAI41111	standard; DNA; 722 BP.
ID	AAI41111	
AC	AAI41111;	
XX		
DT	16-OCT-2002	(first entry)
XX		
DE	GAML related Y-chromosome DNA sequence.	
XX		
KW	Goat embryo sexual identification technique; goat amelogenin gene; GAML;	
KW	sex-specific; gene; ds; Y-chromosome.	
XX		
OS	Capra hircus.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	35..658
FT		/tag= a
XX		/product= "Y-chromosome protein"
PN	TW454013-A.	
XX		
PD	11-SEP-2001.	
XX		
PF	10-NOV-1999;	99TW-00119616.
XX		
PR	10-NOV-1999;	99TW-00119616.
XX		
PA	(CHEN/) CHEN C.	
PA	(JANG/) JANG J.	
PA	(WENG/) WENG T.	
PA	(JENG/) JENG D.	
XX		
PI	Chen C, Jang J, Weng T, Jeng D;	
XX		
DR	WPI; 2002-442016/47.	
XX		
PT	P-PSDB; AAO22534.	
XX		
PT	Sex-specific sequence of goat amelogenin gene, useful for embryo sexual	
PT	identification, comprises high sensitivity even using single white blood	
XX	cell or cleavage c.	
XX		
PS	Disclosure; Page 28; 35pp; Chinese.	
XX		

CC The invention relates to a goat embryo sexual identification technique  
 CC with high efficiency, sensitivity and repeatability. This technique  
 CC involves separately cloning and sequencing the coding regions and the  
 CC introns of the goat amelogenin gene (GAML) on the goat chromosomes. The  
 CC results indicate that there are sex-specific sequences in the fifth  
 CC intron of the gene. The major characteristics according to the present  
 CC invention include high sensitivity, applicable in sex identification even  
 CC only using a single white blood cell or a single cleavage cell of  
 CC blastula; high diagnostic efficiency, capable of identifying hundreds of  
 CC goat embryo in 3 hours; simple operation procedures without complicated  
 CC steps of DNA extraction and need no additional control group intron; and  
 CC can be applied on different species of goats. This polynucleotide  
 CC sequence represents a GAML related Y-chromosome DNA sequence of the  
 CC invention  
 XX

SO Sequence 722 BP, 173 A, 264 C, 140 G, 145 T, 0 U, 0 Other;  
 Query Match 100.0%; Score 22; DB 6; Length 722;  
 Best Local Similarity 100.0%; Pred. No. 3.5;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATTTGTTGGCTGCTCTG 22  
 |||||  
 46 GATTTGTTGGCTGCTCTG 67

Db

RESULT 2  
 AAL41110  
 ID AAL41110 standard; DNA, 752 BP.  
 AC AAL41110;  
 DT 16-OCT-2002 (first entry)  
 XX  
 XX  
 DE GAML related X-chromosome DNA sequence.  
 KW Goat embryo sexual identification technique; goat amelogenin gene; GAML;  
 XX sex-specific; gene; db; X-chromosome.  
 OS Capra hircus.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 35..658  
 FT /tag= a  
 FT /product= "X-chromosome protein"  
 PN TW454013-A.  
 XX  
 PD 11-SEP-2001.  
 XX  
 PF 10-NOV-1999; 99TW-00119616.  
 XX  
 PR 10-NOV-1999; 99TW-00119616.  
 XX  
 PA (CHEN/) CHEN C.  
 PA (JANG/) JANG J.  
 PA (WENG/) WENG T.  
 PA (JENG/) JENG D.  
 XX  
 PI Chen C, Jang J, Weng T, Jeng D;  
 XX  
 DR WPI: 2002-442016/47.  
 DR P-PsDB: MA022534.  
 XX  
 PT Sex-specific sequence of goat amelogenin gene, useful for embryo sexual  
 PT identification, comprises high sensitivity even using single white blood  
 PT cell or cleavage c.  
 XX  
 PS Disclosure: Page 28; 35pp; Chinese.  
 XX  
 CC The invention relates to a goat embryo sexual identification technique  
 CC with high efficiency, sensitivity and repeatability. This technique  
 CC involves separately cloning and sequencing the coding regions and the

CC introns of the goat amelogenin gene (GAML) on the goat chromosomes. The  
 CC results indicate that there are sex-specific sequences in the fifth  
 CC intron of the gene. The major characteristics according to the present  
 CC invention include high sensitivity, applicable in sex identification even  
 CC only using a single white blood cell or a single cleavage cell of  
 CC blastula; high diagnostic efficiency, capable of identifying hundreds of  
 CC goat embryo in 3 hours; simple operation procedures without complicated  
 CC steps of DNA extraction and need no additional control group intron; and  
 CC can be applied on different species of goats. This polynucleotide  
 CC sequence represents a GAML related X-chromosome DNA sequence of the  
 CC invention  
 XX

SO Sequence 752 BP, 184 A, 271 C, 152 G, 145 T, 0 U, 0 Other;  
 Query Match 100.0%; Score 22; DB 6; Length 752;  
 Best Local Similarity 100.0%; Pred. No. 3.5;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATTTGTTGGCTGCTCTG 22  
 |||||  
 46 GATTTGTTGGCTGCTCTG 67

Db

RESULT 3  
 ABN28389  
 ID ABN28389 standard; DNA, 65 BP.  
 AC ABN28389;  
 DT 15-JUL-2002 (first entry)  
 XX  
 XX  
 DE Rat spliced transcript detection oligonucleotide SEQ ID NO:1137.  
 KW Human; mouse; rat; splice transcript; detection; RNA transcript;  
 XX splice variant; transcriptome; oligonucleotide library; ss.  
 OS Rattus norvegicus.  
 PN WO200210449-A2.  
 XX  
 PD 07-FEB-2002.  
 XX  
 PF 20-JUL-2001; 2001WO-IB001903.  
 XX  
 PR 28-JUL-2000; 2000US-0221607P.  
 PR 02-MAY-2001; 2001US-0287724P.  
 XX  
 PA (COMP-) COMPUGEN INC.  
 XX  
 PI Shoshan A, Wasserman A, Mintz E, Mintz L, Faigler S;  
 XX  
 DR WPI: 2002-257383/30.  
 XX  
 PT New oligonucleotide libraries comprising oligonucleotides which  
 PT selectively hybridize to mRNAs transcribed from a transcription unit of a  
 PT genome, useful for detecting tissue-, pathology-, and developmental-  
 PT specific genes.  
 XX  
 PS Example 1; SEQ ID NO 1137; 47pp; English.  
 XX  
 CC The present invention describes oligonucleotide libraries for detecting  
 CC messenger RNAs that populate a (sub-)transcriptome, where the (sub-  
 CC )transcriptome comprises messenger RNAs transcribed from multiple  
 CC transcription units that populate a genome. The library comprises several  
 CC oligonucleotides, each capable of hybridizing selectively to a set of  
 CC messenger RNAs transcribed from a given transcription unit of the genome,  
 CC which encodes one or more messenger RNA splice variants. The  
 CC oligonucleotide libraries are useful for detecting mRNAs from a  
 CC biological sample, in expression profiling studies, in qualitatively or  
 CC quantitatively characterizing the corresponding transcriptome, and in  
 CC detecting RNA transcripts and splice variants of human or animal  
 CC transcriptomes. The libraries may also be used as specialised mini  
 CC libraries to detect transcripts of a sub-transcriptome under a particular



CC biological or pathological state, and so allowing the detection of tissue  
 CC - and pathology-specific genes such as those genes only expressed in  
 CC specific tissue under a specific pathological condition; to detect  
 CC developmental specific genes; and to detect RNA transcripts and splice  
 CC variants of a transcriptome of a patient suffering from a particular  
 CC disorder. ABN27253 to ABN59589 represent oligonucleotide sequences from  
 CC rats, humans and mice, which are used in the exemplification of the  
 CC present invention. N.B. The sequence data for this patent did not form  
 CC part of the printed specification, but was obtained in electronic format  
 CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 SQ Sequence 65 BP; 8 A; 23 C; 14 G; 20 T; 0 U; 0 Other;

Query Match 92.7%; Score 20.4; DB 6; Length 65;  
 Best Local Similarity 95.5%; Pred. No. 14;  
 Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 1 GATTTGTTGGCTGCTCCTG 22  
 6 GATCTTGTTCCTGCTCCTG 27

RESULT 4

ADM80835  
 ID ADM80835 standard; cDNA; 270 BP.

AC ADM80835;

DT 03-JUN-2004 (first entry)

DE Human CADEC-22 encoding cDNA SEQ ID NO:64.

human; cell adhesion and extracellular matrix protein; CADEC-22;  
 neuroprotective; cytoskeletal; anorectic; immune disorder;  
 neurological disorder; developmental disorder;  
 connective tissue disorder; cell proliferative disorder; cancer; obesity;  
 Tangle disease; gene; ss.

OS Homo sapiens.

Key Location/Qualifiers  
 35..212  
 /tag= a  
 /product= "CADEC-22"

FT CDS  
 FT WO2004015396-A2.

PN 19-FEB-2004.

PD 12-AUG-2003; 2003WO-US025418.

PR 13-AUG-2002; 2002US-0403781P.

PR 30-AUG-2002; 2002US-0407034P.

PR 13-SEP-2002; 2002US-0410566P.

PR 24-SEP-2002; 2002US-0413482P.

PR 25-SEP-2002; 2002US-0413890P.

PR 08-NOV-2002; 2002US-0424504P.

PR 13-NOV-2002; 2002US-0426222P.

XX (INCY-) INCYTE CORP.

PI Elliott VS, Khare R, Emerling BM, Kable AE, Tran UK, Jin P;  
 PI Becha SD, Margulis JP, Swarnakar A, Chavla NK, Ramkumar J;  
 PI Hafalia AJA, Lee SY, Jiang X, Jackson AA, Richardson TW, Blake JJ;  
 PI Wang JT, Chien D, Yang YG;

DR MPI: 2004-191795/18.

DR P-PSDB; ADM80793.

XX New cell adhesion and extracellular matrix proteins, useful in  
 PT diagnosing, treating and preventing immune, neurological, developmental,  
 PT connective tissue and cell proliferative disorders including cancer.

PS Claim 5; SEQ ID NO 64; 272pp; English.

CC The present sequence encodes a human cell adhesion and extracellular  
 CC matrix protein designated CADEC-22. CADEC-22 sequences has neuroprotective,  
 CC cytoskeletal and anorectic activities. The CADEC-22 polypeptides and  
 CC polynucleotides are useful in diagnosing, treating and preventing immune,  
 CC neurological, developmental, connective tissue and cell proliferative  
 CC disorders including cancer, e.g. breast, prostate, ovarian, lung or colon  
 CC cancer, obesity and Tangle disease.

SQ Sequence 270 BP; 76 A; 65 C; 67 G; 62 T; 0 U; 0 Other;

Query Match 92.7%; Score 20.4; DB 12; Length 270;  
 Best Local Similarity 95.5%; Pred. No. 16;  
 Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 1 GATTTGTTGGCTGCTCCTG 22  
 47 GATTTATTGCTGCTCCTG 68

RESULT 5

ADM80836  
 ID ADM80836 standard; cDNA; 318 BP.

AC ADM80836;

DT 03-JUN-2004 (first entry)

DE Human CADEC-23 encoding cDNA SEQ ID NO:65.

human; cell adhesion and extracellular matrix protein; CADEC-23;  
 neuroprotective; cytoskeletal; anorectic; immune disorder;  
 neurological disorder; developmental disorder;  
 connective tissue disorder; cell proliferative disorder; cancer; obesity;  
 Tangle disease; gene; ss.

OS Homo sapiens.

Key Location/Qualifiers  
 35..260  
 /tag= a  
 /product= "CADEC-23"

FT CDS  
 FT WO2004015396-A2.

PN 19-FEB-2004.

PD 12-AUG-2003; 2003WO-US025418.

PR 13-AUG-2002; 2002US-0403781P.

PR 30-AUG-2002; 2002US-0407034P.

PR 13-SEP-2002; 2002US-0410566P.

PR 24-SEP-2002; 2002US-0413482P.

PR 25-SEP-2002; 2002US-0413890P.

PR 08-NOV-2002; 2002US-0424504P.

PR 13-NOV-2002; 2002US-0426222P.

XX (INCY-) INCYTE CORP.

PI Elliott VS, Khare R, Emerling BM, Kable AE, Tran UK, Jin P;  
 PI Becha SD, Margulis JP, Swarnakar A, Chavla NK, Ramkumar J;  
 PI Hafalia AJA, Lee SY, Jiang X, Jackson AA, Richardson TW, Blake JJ;  
 PI Wang JT, Chien D, Yang YG;

DR MPI: 2004-191795/18.

DR P-PSDB; ADM80794.

XX New cell adhesion and extracellular matrix proteins, useful in  
 PT diagnosing, treating and preventing immune, neurological, developmental,  
 PT connective tissue and cell proliferative disorders including cancer.

PS Claim 5; SEQ ID NO 65; 272pp; English.

XX The present sequence encodes a human cell adhesion and extracellular  
CC matrix protein designated CADECM. CADECM sequences has neuroprotective,  
CC cyostatic and anorectic activities. The CADECM polypeptides and  
CC polynucleotides are useful in diagnosing, treating and preventing immune,  
CC neurological, developmental, connective tissue and cell proliferative  
CC disorders including cancer, e.g. breast, prostate, ovarian, lung or colon  
CC cancer, obesity and Tangier disease.  
XX  
SQ Sequence 318 BP, 87 A; 81 C; 75 G; 75 T; 0 U; 0 Other;

Query Match 92.7%; Score 20.4; DB 12; Length 318;  
Best Local Similarity 95.5%; Pred. No. 17;  
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GATTTGTTGGCCCTCCTCG 22  
|||||  
Db 47 GATTTGTTGGCCCTCCTCG 68

RESULT 6  
AAZ50832  
ID AAZ50832 standard; DNA; 476 BP.

XX  
AC AAZ50832;  
XX  
DT 31-MAY-2000 (first entry)  
XX  
DE Rat amelogenin gene (A4).  
XX  
KM Amelogenin; splice variant; rat; (A4); chondrogenesis; osteogenesis;  
KM chondrogenic inducing molecule; CIM; cartilage growth; osteopathic;  
KM extracellular matrix protein; tooth enamel; enamel mineralisation;  
KM ameloblast; bone regeneration; composite cell construct; ds.  
XX  
OS Rattus sp.

XX  
FH Key

FT exon Location/Qualifiers  
1..36  
/\*tag= a

FT exon /partial  
/number= 1  
37..101  
/\*tag= b

FT exon /number= 2  
48..317

FT CDS  
/\*tag= c  
/product= "Rat amelogenin protein"  
/transl\_except= (pos:69..71, aa:Gly)

FT sig\_peptide  
48..95  
/\*tag= d

FT mat\_peptide  
96..314  
/\*tag= e  
/label= Mature\_rat\_amelogenin\_protein

FT exon  
102..149  
/\*tag= f

FT exon /number= 3  
150..191  
/\*tag= g

FT exon /number= 4  
192..236  
/\*tag= h

FT exon /number= 5  
237..311  
/\*tag= i

FT exon /number= 6d  
/note= "Comprises of gene segments 6a, b, c and d"  
312..317  
/\*tag= j  
/number= 7  
/note= "Includes the stop codon"

FT exon  
XX  
PN WO200006734-A1.

XX  
PD 10-FEB-2000.

XX 29-JUL-1999; 99WO-US017342.

XX 29-JUL-1998; 98US-0094489P.

XX (NOUN) UNIV NORTHWESTERN.

XX Vets A, Nebgen DR;

XX WPI; 2000-205464/18.

XX P-PBDB; AA145074.

XX Novel amelogenin polypeptides and polynucleotides, useful for enhancing  
PT bone generation in mammals and synthesizing bone matrix or articular  
PT surfaces at implant sites.

PS Example 2; Fig 11B; 79pp; English.

XX The present DNA sequence is the full-length rat amelogenin gene (A4),  
CC comprising exons 1-7, including the exon segment 6d. It is derived from  
CC the rat incisor odontoblast-pulp cDNA library. The splice variants of  
CC this gene functions as an osteogenic or chondrogenic inducing molecule  
CC (CIM), which is useful for enhancing bone or cartilage growth. It has  
CC osteopathic activity. Amelogenin belongs to the family of extracellular  
CC matrix proteins, in developing tooth enamel, that are produced by the  
CC ameloblasts and plays a role in enamel mineralisation. Chondrogenic or  
CC osteogenic inducing amelogenin molecules are useful to induce  
CC differentiation of cells to the osteogenic and chondrogenic phenotypes  
CC and can be used in a composite cell construct for bone and cartilage  
CC regeneration. The polynucleotides can be employed to produce the  
XX polypeptides by recombinant techniques

SQ Sequence 476 BP; 155 A; 106 C; 98 G; 117 T; 0 U; 0 Other;

Query Match 92.7%; Score 20.4; DB 3; Length 476;  
Best Local Similarity 95.5%; Pred. No. 18;  
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GATTTGTTGGCCCTCCTCG 22  
|||||  
Db 59 GATCTGTTGCTCCTCCTCG 80

RESULT 7

ADM80826  
ID ADM80826 standard; cDNA; 549 BP.

AC ADM80826;

DT 03-JUN-2004 (first entry)

XX Human CADECM-13 encoding cDNA SEQ ID NO:55.

XX human; cell adhesion and extracellular matrix protein; CADECM;  
KW neuroprotective; cyostatic; anorectic; immune disorder;  
KW neurological disorder; developmental disorder;  
KW connective tissue disorder; cell proliferative disorder; cancer; obesity;  
KW Tangier disease; gene; ss.

XX Homo sapiens.

OS Key Location/Qualifiers  
35..491  
/\*tag= a  
/product= "CADECM-13"

FT CDS  
WO2004015396-A2.

XX 19-FEB-2004.

XX 12-AUG-2003; 2003WO-US025418.

```
XX 13-AUG-2002; 2002US-0403781P.
PR 30-AUG-2002; 2002US-0407034P.
PR 13-SEP-2002; 2002US-0410566P.
PR 24-SEP-2002; 2002US-0413482P.
PR 25-SEP-2002; 2002US-0413890P.
PR 08-NOV-2002; 2002US-0424904P.
PR 13-NOV-2002; 2002US-0426222P.
XX (INCY-) INCYTE CORP.
PA
XX Eliott VS, Khare R, Emerling BM, Kable AE, Tran UK, Jin P,
PI Becha SD, Marquis JP, Swarnakar A, Chawla NK, Ramkumar J,
PI Hafalla AJD, Lee SY, Jiang X, Jackson AA, Richardson TW, Blake JJ,
PI Wang JT, Chien D, Yang YG;
DR WPI; 2004-191795/18.
DR P-PEDB; ADM80784.
XX
XX New cell adhesion and extracellular matrix proteins, useful in
PT diagnosing, treating and preventing immune, neurological, developmental,
PT connective tissue and cell proliferative disorders including cancer.
XX Claim 5; SEQ ID NO 55; 272pp; English.
XX
XX The present sequence encodes a human cell adhesion and extracellular
CC matrix protein designated CADPCM. CADPCM sequences has neuroprotective,
CC cytoskeletal and anorectic activities. The CADPCM polypeptides and
CC polynucleotides are useful in diagnosing, treating and preventing immune,
CC neurological, developmental, connective tissue and cell proliferative
CC disorders including cancer, e.g. breast, prostate, ovarian, lung or colon
CC cancer, obesity and Tangle disease.
XX
SQ Sequence 549 BP, 134 A; 203 C; 112 G; 100 T; 0 U; 0 Other;
Query Match 92.7%; Score 20.4; DB 12; Length 549;
Best Local Similarity 95.5%; Pred. No. 18;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 GATTTGTTGGCTGCTCTG 22
Db 47 GATTTATTTGCTGCTCTG 68
RESULT 8
ADBS9026
ID ADBS9026 standard; DNA; 556 BP.
XX
AC ADBS9026;
XX
XX 04-DEC-2003 (first entry)
DT
XX
XX Toxicity-related gene, SEQ ID 4052.
DE
XX
XX Toxic; toxin; gene expression profile; hepatotoxicity; liver;
KW drug screening; toxicity assay; ds.
XX
XX Unidentified.
OS
XX WO2003064624-A2.
XX
XX 07-AUG-2003.
PD
XX
XX 31-JAN-2003; 2003WO-US003194.
PF
XX
XX 31-JAN-2002; 2002US-00060087.
PR 15-MAR-2002; 2002US-0364045P.
PR 15-MAR-2002; 2002US-0364055P.
PR 30-DEC-2002; 2002US-0436643P.
XX
XX (GENE-) GENE LOGIC INC.
PA
XX Mendrick D, Porter M, Johnson K, Higge B, Castle A, Elashoff M;
```

```
XX WPI; 2003-689530/65.
DR
XX Predicting a toxic effect of a compound, useful in identifying toxicity
XX markers in liver tissues or cells for drug screening and toxicity assays,
PT comprises preparing gene expression profile of tissue or cells exposed to
PT the compound.
XX Claim 1; SEQ ID NO 4052; 1156pp; English.
XX
XX The present invention relates to a method for predicting a toxic effect
CC of a compound. The method comprises preparing a gene expression profile
CC of a tissue or cell sample exposed to the compound, and comparing the
CC gene expression profile to a database comprising SEQ ID 1-4925, where
CC differential expression of the gene indicates at least one toxic effect.
CC The method is useful for predicting at least one toxic effect of a
CC compound, predicting hepatotoxicity or the progression of a toxic effect
CC of a compound, identifying an agent that modulates the onset or
CC progression of a toxic response, predicting the cellular pathways that a
CC compound modulates in a cell, and identifying an agent that modulates at
CC least one activity of a protein. The method and compositions of the
CC present invention using a database of genes having liver toxin-induced
CC differential expression, are useful in identifying toxicity markers in
CC liver tissues or cells for drug screening and toxicity assays. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 556 BP; 121 A; 183 C; 114 G; 119 T; 0 U; 19 Other;
Query Match 92.7%; Score 20.4; DB 10; Length 556;
Best Local Similarity 95.5%; Pred. No. 18;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 GATTTGTTGGCTGCTCTG 22
Db 33 GATCTTGTGCTGCTCTG 54
RESULT 9
ADBS3782
ID ADBS3782 standard; DNA; 556 BP.
XX
XX ADBS3782;
XX
XX 04-DEC-2003 (first entry)
DT
XX
XX Primary rat hepatocyte toxicity modelling related gene SEQ ID NO:4324.
DE
XX
XX toxic effect; gene expression profile; hepatotoxicity; diagnostic marker;
KW toxicity marker; toxicity progression; drug screening;
KW primary rat hepatocyte toxicity modelling; gene; ds.
XX
XX Rattus norvegicus.
OS
XX WO2003065993-A2.
XX
XX 14-AUG-2003.
PD
XX
XX 04-FEB-2003; 2003WO-US003482.
PF
XX
XX 04-FEB-2002; 2002US-0353171P.
PR 13-MAR-2002; 2002US-036534P.
PR 08-APR-2002; 2002US-0370248P.
PR 10-APR-2002; 2002US-0371134P.
PR 10-APR-2002; 2002US-0371135P.
PR 10-APR-2002; 2002US-0371150P.
PR 11-APR-2002; 2002US-0371413P.
PR 11-APR-2002; 2002US-0373601P.
PR 19-APR-2002; 2002US-0374139P.
PR 22-APR-2002; 2002US-0378370P.
PR 08-MAY-2002; 2002US-0378652P.
PR 09-MAY-2002; 2002US-0378652P.
```

PR 09-MAY-2002; 2002US-0378653P.  
PR 09-MAY-2002; 2002US-0378655P.  
PR 09-JUL-2002; 2002US-0394233P.  
PR 09-JUL-2002; 2002US-0394235P.  
PR 04-SEP-2002; 2002US-0407688P.  
PR 28-JAN-2003; 2003US-0442900P.  
PA (GENE-) GENE LOGIC INC.  
XX  
PI Mendrick D, Potter M, Johnson K, Higgs B, Caetle A, Orr M,  
PI Elshoff M;  
XX  
DR WPI; 2003-731472/69.  
XX  
XX  
PT Determining if a compound induces a toxic effect on a tissue or cell, for  
PT identifying hepatotoxic compounds, comprises comparing a gene expression  
PT profile of a tissue or cell sample to a database of Tox mean and non-Tox  
PT mean values.  
XX  
PS Claim 44; SEQ ID NO 4324; 874bp; English.  
XX  
XX The present invention describes a method for determining whether a  
CC compound induces a toxic effect on a tissue or cell. The method comprises  
CC preparing a gene expression profile of a tissue or cell sample exposed to  
CC the compound, and comparing the gene expression profile to a database  
CC comprising data or information on the Tox mean and non-Tox mean value.  
CC The method is useful for predicting or identifying at least one toxic  
CC effect, particularly hepatotoxicity, of a test or unknown compound. The  
CC genes listed in the specification are useful as diagnostic or toxicity  
CC markers for the prediction or identification of the physiological state  
CC of tissue or cell sample that has been exposed to a compound, or to  
CC identify or predict the toxic effects of a compound or an agent. These  
CC may also be used as markers for monitoring toxicity progression or for  
CC drug screening. The present sequence represents a primary rat hepatocyte  
CC toxicity modelling related gene sequence from the present invention.  
XX  
SQ Sequence 556 BP; 121 A; 183 C; 114 G; 119 T; 0 U; 19 Other;  
XX  
Query Match 92.7%; Score 20.4; DB 10; Length 556;  
Best Local Similarity 95.5%; Pred. No. 18;  
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 GATTTGTTGCTGCTGCTG 22  
DB 33 GATCTGTGCTGCTGCTG 54  
XX  
RESULT 10  
ADM80827  
ID ADM80827 standard; cDNA; 623 BP.  
XX  
AC ADM80827;  
XX  
DT 03-JUN-2004 (first entry)  
XX  
DE Human CADECM-14 encoding cDNA SEQ ID NO:56.  
XX  
KW human; cell adhesion and extracellular matrix protein; CADECM;  
KW neuroprotective; cytostatic; anorectic; immune disorder;  
KW neurological disorder; developmental disorder;  
KW connective tissue disorder; cell proliferative disorder; cancer; obesity;  
KW Tangier disease; gene; ss.  
OS  
XX Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT CDS 36..563  
FT /\*tag= a  
FT /product= "CADECM-14"  
XX  
XX WO2004015396-A2.  
XX  
PD 19-FEB-2004.

XX  
XX 12-AUG-2003; 2003WO-US025418.  
XX  
PR 13-AUG-2002; 2002US-0403781P.  
PR 10-AUG-2002; 2002US-0407034P.  
PR 13-SEP-2002; 2002US-0410566P.  
PR 24-SEP-2002; 2002US-0413482P.  
PR 25-SEP-2002; 2002US-0413890P.  
PR 08-NOV-2002; 2002US-0424904P.  
PR 13-NOV-2002; 2002US-0426222P.  
XX  
PA (INCY-) INCYTE CORP.  
XX  
PI Elliott VS, Khare R, Emerling BM, Kable AE, Tran UK, Jin P;  
PI Becha SD, Margulis JP, Swarnakar A, Chawla NK, Ramkumar J;  
PI Hafalia AYA, Lee SY, Jiang X, Jackson AA, Richardson TW, Blake JT;  
PI Wang JT, Chien D, Yang YG;  
XX  
DR WPI; 2004-191795/18.  
DR P-PSDB; ADM80785.  
XX  
XX New cell adhesion and extracellular matrix proteins, useful in  
PT diagnosing, treating and preventing immune, neurological, developmental,  
PT connective tissue and cell proliferative disorders including cancer.  
XX  
PS Claim 5; SEQ ID NO 56; 272bp; English.  
XX  
XX The present sequence encodes a human cell adhesion and extracellular  
CC matrix protein designated CADECM. CADECM sequences has neuroprotective,  
CC cyostatic and anorectic activities. The CADECM polypeptides and  
CC polynucleotides are useful in diagnosing, treating and preventing immune,  
CC neurological, developmental, connective tissue and cell proliferative  
CC disorders including cancer; e.g. breast, prostate, ovarian, lung or colon  
CC cancer, obesity and Tangier disease.  
XX  
SQ Sequence 623 BP; 150 A; 228 C; 128 G; 117 T; 0 U; 0 Other;  
XX  
Query Match 92.7%; Score 20.4; DB 12; Length 623;  
Best Local Similarity 95.5%; Pred. No. 18;  
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 GATTTGTTGCTGCTGCTG 22  
DB 47 GATTTATTGCTGCTGCTG 68  
XX  
RESULT 11  
AAZ07020  
ID AAZ07020 standard; DNA; 750 BP.  
XX  
AC AAZ07020;  
XX  
DT 15-NOV-1999 (first entry)  
XX  
DE Amelogenin X nucleotide sequence.  
XX  
KW Amelogenin X; Amelogenin Y; genotype; 5' nuclease amplification;  
KW multiple allelic sites; apolipoprotein E; apoE; coronary artery disease;  
KW Alzheimer's disease; ds.  
OS  
XX Unidentified.  
XX  
XX WO9940226-A2.  
XX  
XX 12-AUG-1999.  
XX  
PR 08-JAN-1999; 99WO-US000499.  
XX  
PR 04-FEB-1998; 98US-00018595.  
XX  
PA (PEKE ) PERKIN-ELMER CORP.  
XX  
PI Liivak KI, Goodsaiaid F;

```
XX
DR 1999-539985/45.
XX
PT 5' nuclease amplification assay using fluorescence-quencher probes for
XX determination of a genotype at multiple allelic sites.
XX
PS Disclosure; Fig 10; 95pp; English.
XX
CC The present invention describes first and second sets of fluorescer-
CC quencher probes used simultaneously in a 5' nuclease assay to identify
CC which members of a first or second set of substantially homologous
CC sequences are present in a DNA sample. The method can be used to genotype
CC a sample of genomic DNA at two or more different allelic sites.
CC
CC Generating a fluorescence spectrum and signature for each genotype, which
CC uniquely reflects the assay's inherent inefficiency for that genotype
CC given the particular conditions, probes and primers used, the genotype of
CC unknown sequences can be determined. The assay was shown to be useful for
CC determining apog genotypes. The assay can be used as a diagnostic tool
CC for assessing the risk for coronary artery disease and/or late-onset
CC Alzheimer's disease. Using the 5' nuclease assay of the invention it is
CC possible to determine a genotype at two or more allelic sites in a single
CC reaction. This approach is much faster than previous approaches to
CC genotyping genes having two or more allelic sites, such as the
CC apolipoprotein E gene. A key advantage of the method for determining the
CC genotype of a sample of DNA at multiple allelic sites is that it does not
CC rely on 5' nuclease assay working with 100% efficiency to distinguish
CC between substantially homologous sequences such as alleles. The present
CC sequence represent the nucleotide sequence for ameloogenin X, which is
CC used in the exemplification of the present invention
XX
SQ Sequence 750 BP; 191 A; 260 C; 146 G; 153 T; 0 U; 0 Other;
XX
Query Match 92.7%; Score 20.4; DB 2; Length 750;
Best Local Similarity 95.5%; Pred. No. 19;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
QY 1 GATTGTGCTGCTGCTGCTG 22
DB 80 GATTGTGCTGCTGCTGCTG 101
XX
RESULT 12
AAZ07018
ID AAZ07018 standard; DNA; 793 BP.
XX
AC AAZ07018;
XX
DT 15-NOV-1999 (first entry)
XX
DE Ameloogenin X nucleotide sequence.
XX
KW Ameloogenin X; Ameloogenin Y; genotype; 5' nuclease amplification;
XX multiple allelic site; apolipoprotein E; apog; coronary artery disease;
XX Alzheimer's disease; ss.
XX
OS Unidentified.
XX
PN WO940226-A2.
XX
PD 12-AUG-1999.
XX
PF 08-JAN-1999; 99WO-US000499.
XX
PR 04-FEB-1998; 98US-00018595.
XX
PA (PERKE ) PERKIN-ELMER CORP.
XX
PI Llyak KJ, Goodsaiaid F;
XX
DR WPI; 1999-539985/45.
XX
PT 5' nuclease amplification assay using fluorescence-quencher probes for
XX determination of a genotype at multiple allelic sites.
```

```
XX
PS Disclosure; Fig 8A; 95pp; English.
XX
CC The present invention describes first and second sets of fluorescer-
CC quencher probes used simultaneously in a 5' nuclease assay to identify
CC which members of a first or second set of substantially homologous
CC sequences are present in a DNA sample. The method can be used to genotype
CC a sample of genomic DNA at two or more different allelic sites.
CC
CC Generating a fluorescence spectrum and signature for each genotype, which
CC uniquely reflects the assay's inherent inefficiency for that genotype
CC given the particular conditions, probes and primers used, the genotype of
CC unknown sequences can be determined. The assay was shown to be useful for
CC determining apog genotypes. The assay can be used as a diagnostic tool
CC for assessing the risk for coronary artery disease and/or late-onset
CC Alzheimer's disease. Using the 5' nuclease assay of the invention it is
CC possible to determine a genotype at two or more allelic sites in a single
CC reaction. This approach is much faster than previous approaches to
CC genotyping genes having two or more allelic sites, such as the
CC apolipoprotein E gene. A key advantage of the method for determining the
CC genotype of a sample of DNA at multiple allelic sites is that it does not
CC rely on 5' nuclease assay working with 100% efficiency to distinguish
CC between substantially homologous sequences such as alleles. The present
CC sequence represent the nucleotide sequence for ameloogenin X, which is
CC used in the exemplification of the present invention
XX
SQ Sequence 793 BP; 215 A; 266 C; 150 G; 162 T; 0 U; 0 Other;
XX
Query Match 92.7%; Score 20.4; DB 2; Length 793;
Best Local Similarity 95.5%; Pred. No. 19;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
QY 1 GATTGTGCTGCTGCTGCTG 22
DB 80 GATTGTGCTGCTGCTGCTG 101
XX
RESULT 13
ADQ18492
ID ADQ18492 standard; DNA; 259202 BP.
XX
AC ADQ18492;
XX
DT 26-AUG-2004 (first entry)
XX
DE Human soft tissue sarcoma-upregulated DNA - SEQ ID 1311.
XX
KW soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;
XX da.
XX
OS Homo sapiens.
XX
PN WO2004048938-A2.
XX
PD 10-JUN-2004.
XX
PF 26-NOV-2003; 2003WO-US038193.
XX
PR 26-NOV-2002; 2002US-0429739P.
XX
PA (PROT-) PROTEIN DESIGN LABS INC.
XX
PI Aziz N, Ginsburg WM, Zlotnick A;
XX
DR WPI; 2004-441208/41.
XX
PT Early detection of soft tissue sarcoma comprises determining expression
XX of a gene in a first soft tissue sample and a normal soft tissue sample
XX and comparing the gene expression, also useful in treating soft tissue
XX sarcoma.
XX
PS Example 2; SEQ ID NO 1311; 210pp; English.
XX
PT The invention relates to a novel method for detecting soft tissue sarcoma
```

CC which comprises obtaining a first soft tissue sample from an individual  
CC and a normal soft tissue sample from the same or different individual,  
CC determining the expression of a gene in both samples and comparing the  
CC expression of the gene in both soft tissue samples, where a higher level  
CC of protein expression in the first soft tissue sample indicates the  
CC presence of soft tissue sarcoma. The method of the invention has  
CC cytostatic applications and may be useful for detecting soft tissue  
CC sarcoma, possibly via gene therapy or vaccine production. The nucleic  
CC acid sequences may be useful in diagnostic and screening applications.  
CC The current sequence is that of a human soft tissue sarcoma-upregulated  
CC DNA of the invention. The current sequence is not shown within the  
CC specification per se but was submitted in CD format by the inventor.

SQ Sequence 259202 BP; 81659 A; 51421 G; 49221 G; 76861 T; 0 U; 0 Other;

Query Match 92.7%; Score 20.4; DB 12; Length 259202;  
Best Local Similarity 95.5%; Pred. No. 39;  
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Dy 1 GATTTGTTGCTGCTGCTGCTG 22  
Db 69884 GATTTATTTGCTGCTGCTGCTG 69905

RESULT 14  
AA18413  
ID AA18413 standard; DNA; 20 BP.

AC AA18413;

DT 11-MAY-1999 (first entry)

DE PCR primer BE2(+) for bovine amelogenin gene.

KW bovine embryo sexing; PCR primer; 59;  
KW bovine embryo sexing; PCR primer; 59.

OS Synthetic.  
OS Bos sp.

PN US5876942-A.

PD 02-MAR-1999.

PP 24-JUL-1997; 97US-00899811.

PR 24-JUL-1997; 97US-00899811.

PA (NASC-) NAT SCI COUNCIL REPUBLIC OF CHINA.

PI Chao K, Wang C, Cheng WT, Chen C, Hu C;

DR WPI; 1999-189629/16.

PT New oligonucleotide primers based on bovine amelogenin gene, intron 5  
PT sequences - useful for sexing cows by Polymerase Chain Reaction studies.

PS Disclosure; Col 6; 28pp; English.

CC This sequence is a PCR primer for the Holstein cow amelogenin (BAML) gene  
CC The invention relates to an oligonucleotide primer set, useful for bovine  
CC embryo sexing, that comprises two primers, each of which can hybridise  
CC specifically and simultaneously, to an intron 5 sequence of BAML, located  
CC on the bovine X and Y chromosomes. The primers may be used in a rapid,  
CC highly reproducible and sensitive method for determining the sex of  
CC bovine embryos, which involves PCR of the BAML genes located on the X and  
CC Y chromosomes of Holstein dairy cattle. In order to use PCR in sex-  
CC determination studies, a nucleotide sequence, specific against sex,  
CC has to be produced (e.g. one associated with testis determining factor).  
CC However, in this PCR based method, each primer can only recognise DNA  
CC fragments from one, not both, of the sex chromosomes, therefore, internal  
CC control primers, derived from the subject gene have to be added to the  
CC reaction. This can result in competition between the primers, or the

CC formation of dimer primers during amplification, rendering the results  
CC inaccurate. The primers overcome this problem, as they are homologous to  
CC both the X and Y chromosomes, and so amplify DNA from both chromosomes  
CC simultaneously, allowing gender to be determined by quick, simple and  
CC accurate PCR and electrophoresis

SQ Sequence 20 BP; 1 A; 6 C; 3 G; 10 T; 0 U; 0 Other;

Query Match 90.9%; Score 20; DB 2; Length 20;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Dy 2 ATTTGTTGCTGCTGCTGCT 21  
Db 1 ATTTGTTGCTGCTGCTGCT 20

RESULT 15  
ABK92154  
ID ABK92154 standard; DNA; 2190 BP.

AC ABK92154;

DT 15-AUG-2002 (first entry)

DE Prostate cancer-associated DNA sequence #40.

KW Prostate cancer; prostate tumour tissue; human; mammal; cytostatic;  
KW gene therapy; gene; ds.

OS Mammalia.

PN WO200230268-A2.

PD 18-APR-2002.

PP 12-OCT-2001; 2001WO-US032045.

PR 13-OCT-2000; 2000US-00687576.

PR 08-DEC-2000; 2000US-00733288.

PR 24-DEC-2000; 2000US-00733742.

PR 16-JAN-2001; 2001US-0263957P.

PR 16-MAR-2001; 2001US-0276791P.

PR 16-MAR-2001; 2001US-0276888P.

PR 06-APR-2001; 2001US-0281922P.

PR 24-APR-2001; 2001US-0286214P.

PR 30-APR-2001; 2001US-02867046.

PR 04-MAY-2001; 2001US-0288589P.

PA (EOSB-) EOS BIOTECHNOLOGY INC.

PI Gish KC, Mack DH, Wilson KE, Afar D, Hevezl P;

DR WPI; 2002-471335/50.

DR P-PSDB; ABG61839.

PT Detecting a prostate cancer-associated transcript in a cell in a patient,  
PT by determining if prostate cancer-associated genes are expressed in a  
PT prostate tissue.

PS Claim 22; Page 331; 436pp; English.

CC The present invention relates to methods of detecting a prostate cancer-  
CC associated transcript in a cell from a patient. The method comprises  
CC contacting a biological sample from the patient with prostate cancer-  
CC associated polynucleotides (designated PC genes) that selectively  
CC hybridise to a sequence that is at least 80% identical to them. The  
CC prostate cancer-associated polynucleotide sequences are differentially  
CC expressed in prostate tumour tissue or in prostate cancer and are derived  
CC from the tissues of various organisms such as humans or other mammals  
CC (e.g. mice, sheep and dogs). The methods of the invention are useful for  
CC diagnosing and treating prostate cancer in mammals. The prostate cancer-

CC associated genes are useful for diagnosing or treating prostate cancer,  
 CC as well as for identifying modulators of prostate cancer or agents that  
 CC inhibit prostate cancer. The nucleic acid sequences are particularly  
 CC useful in gene therapy, as a vaccine or in antisense applications.  
 CC ABK92115-ABK92263 represent prostate cancer-associated polynucleotide  
 CC sequences

XX  
 SQ Sequence 2190 BP; 583 A; 488 C; 494 G; 625 T; 0 U; 0 Other;

Query Match 88.2%; Score 19.4; DB 6; Length 2190;  
 Best Local Similarity 95.2%; Pred. No. 60;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 ATTTGTTGGCTGCTCTCG 22  
 |||||||  
 Db 2026 ATTTGTTTCTGCTCTCG 2046

Search completed: February 8, 2005, 13:56:51  
 Job time : 35.1839 secs

**This Page Blank (uspto)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using SW model

Run on: February 8, 2005, 13:03:55 : Search time 5.94179 Seconds

(without alignment)  
2631.757 Million cell updates/sec

Title: US-10-754-437-3

Sequence: 1 gatctgtgtgctgctgctctg 22

Scoring table: IDENTITY\_NTC

Gapop 10.0 , Gapext 1.0

Searched: 824507 seqs, 355394441 residues

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

Issued Patents NA: \*  
1: /cgn2\_6/prodata/1/ina/5A\_COMB.seq: \*  
2: /cgn2\_6/prodata/1/ina/5B\_COMB.seq: \*  
3: /cgn2\_6/prodata/1/ina/6A\_COMB.seq: \*  
4: /cgn2\_6/prodata/1/ina/6B\_COMB.seq: \*  
5: /cgn2\_6/prodata/1/ina/PCTUS\_COMB.seq: \*  
6: /cgn2\_6/prodata/1/ina/backfileseq1.seq: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	20.4	92.7	476	4	US-09-744-128-3	Sequence 3, Appl1
2	20.4	92.7	500	2	US-09-018-595B-3	Sequence 3, Appl1
3	20.4	92.7	500	3	US-09-324-709A-1	Sequence 3, Appl1
4	20.4	92.7	793	2	US-018-595B-1	Sequence 1, Appl1
5	20.4	92.7	793	3	US-09-324-709A-1	Sequence 5, Appl1
6	20.4	90.9	20	2	US-08-899-811-5	Sequence 1, Appl1
7	19.4	88.2	4181	4	US-09-976-594-366	Sequence 366, App
8	19.4	86.4	19	2	US-08-899-811-4	Sequence 4, Appl1
9	18.8	85.5	802	2	US-09-018-595B-2	Sequence 2, Appl1
10	18.8	85.5	802	3	US-09-324-709A-2	Sequence 3, Appl1
11	17.8	80.9	173	4	US-09-513-999C-35462	Sequence 35462, A
12	17.4	79.1	116	3	US-09-338-907-110	Sequence 110, App
13	17.4	79.1	116	3	US-09-218-207-110	Sequence 1, App
14	17.4	79.1	5326	3	US-09-338-907-124	Sequence 124, App
15	17.4	79.1	5326	3	US-09-218-207-124	Sequence 124, App
16	17.4	79.1	56516	2	US-08-996-106-1	Sequence 1, Appl1
17	17.4	79.1	56516	3	US-09-338-907-1	Sequence 1, Appl1
18	17.4	79.1	56516	3	US-09-218-207-1	Sequence 1, Appl1
19	17.4	79.1	56520	3	US-09-338-907-179	Sequence 179, App
20	17.4	79.1	56520	3	US-09-218-207-179	Sequence 179, App
21	16.8	76.4	40	4	US-09-313-221A-40	Sequence 40, Appl1
22	16.8	76.4	2734	4	US-09-461-657B-13	Sequence 13, Appl1
23	16.8	76.4	2734	4	US-09-461-657B-11	Sequence 11, Appl1
24	16.8	76.4	2912	4	US-09-461-657B-12	Sequence 12, Appl1
25	16.8	76.4	2987	3	US-07-757-342D-1	Sequence 1, Appl1
26	16.8	76.4	2987	4	US-09-461-657B-1	Sequence 1, Appl1
27	16.8	76.4	3018	4	US-09-016-434-1468	Sequence 1468, Ap

C	28	16.4	74.5	2438	6	5432081-1	Patent No. 5432081
	29	16.2	73.6	433	4	US-09-621-976-13460	Sequence 13460, A
	30	16.2	73.6	586	4	US-09-669-751-55	Sequence 55, Appl1
C	31	16.2	73.6	2148	3	US-09-219-983A-19	Sequence 19, Appl1
C	32	16.2	73.6	2492	3	US-09-219-983A-18	Sequence 18, Appl1
C	33	16.2	73.6	5852	4	US-09-853-768-10	Sequence 10, Appl1
C	34	16.2	73.6	7037	4	US-09-853-768-3	Sequence 3, Appl1
C	35	15.8	71.8	282	4	US-09-248-796A-7539	Sequence 7539, Ap
	36	15.8	71.8	619	4	US-09-270-767-12644	Sequence 12644, A
	37	15.8	71.8	1407	3	US-09-505-250-2	Sequence 2, Appl1
C	38	15.8	71.8	1917	3	US-08-808-346-1	Sequence 1, Appl1
C	39	15.8	71.8	1926	1	US-07-901-703-12	Sequence 12, Appl1
C	40	15.8	71.8	1926	1	US-08-147-023-26	Sequence 26, Appl1
C	41	15.8	71.8	1926	1	US-08-278-729A-22	Sequence 22, Appl1
C	42	15.8	71.8	1926	1	US-08-480-528A-9	Sequence 9, Appl1
C	43	15.8	71.8	1926	1	US-08-479-666-9	Sequence 9, Appl1
C	44	15.8	71.8	1926	1	US-08-155-343A-22	Sequence 22, Appl1
C	45	15.8	71.8	1926	1	US-08-406-672-22	Sequence 22, Appl1

#### ALIGNMENTS

```
RESULT 1
US-09-744-128-3
; Sequence 3, Application US/09744128
; Patent No. 6677306
; GENERAL INFORMATION:
; APPLICANT: Vels et al.
; TITLE OF INVENTION: Chondrogenic and Osteogenic Inducing Molecule
; FILE REFERENCE: 27636/36983
; CURRENT APPLICATION NUMBER: US/09/744,128
; PRIOR FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: PCT/US99/17342
; PRIOR FILING DATE: 1999-07-29
; PRIOR APPLICATION NUMBER: 60/094,489
; PRIOR FILING DATE: 1998-07-29
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Patentln 3.1
; SEQ ID NO 3
; LENGTH: 476
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: PCR Product
US-09-744-128-3

Query Match          92.7%; Score 20.4; DB 4; Length 476;
Best Local Similarity 95.5%; Pred. No. 1.3;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 GATTTGTTGCTGCTGCTG 22
      ||| ||||| ||||| |||||
Db      59 GATCTGTGCTGCTGCTG 80

RESULT 2
US-09-018-595B-3
; Sequence 3, Application US/09018595B
; Patent No. 5962233
; GENERAL INFORMATION:
; APPLICANT: Perkin-Elmer Corporation,
; APPLICANT: Applied Biosystems Division
; TITLE OF INVENTION: DETERMINATION OF GENOTYPE OF
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSER: David J. Weitz,
; ADDRESSER: Wilson Sorensen Goodrich & Rosati
; STREET: 650 Page Mill Road
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
```

```

;
; ZIP: 94304-1050
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch diskette
;
; COMPUTER: IBM compatible
; OPERATING SYSTEM: Microsoft Windows 95
; SOFTWARE: Wordperfect for windows 6.0,
; SOFTWARE: ASCII (DOS) TEXT format
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/018,595B
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: David J. Weitz
; REGISTRATION NUMBER: 38,362
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 493-9300
; TELEFAX: (650) 493-6811
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 500 nucleotides
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
; US-09-018-595B-3
;
Query Match          92.7%; Score 20.4; DB 2; Length 500;
Best Local Similarity 95.5%; Pred. No. 1.4;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GATTGTTGCTGCTGCTCTG 22
Db 80 GATTATTGCTGCTGCTCTG 101

RESULT 3
US-09-324-709A-3
; Sequence 3, Application US/09324709A
; Patent No. 6154707
; GENERAL INFORMATION:
; APPLICANT: Perkin-Elmer Corporation,
; TITLE OF INVENTION: FLUORESCENCE GENOTYPING AT MULTIPLE ALLELIC
; TITLE OF INVENTION: SITES
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David J. Weitz,
; STREET: 650 Page Mill Road
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1050
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch diskette
; OPERATING SYSTEM: Microsoft Windows 95
; SOFTWARE: Wordperfect for windows 6.0,
; SOFTWARE: ASCII (DOS) TEXT format
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/324,709A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: David J. Weitz
; REGISTRATION NUMBER: 38,362
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 493-9300
; TELEFAX: (650) 493-6811
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 793 nucleotides
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
; US-09-018-595B-1
;
Query Match          92.7%; Score 20.4; DB 2; Length 793;
Best Local Similarity 95.5%; Pred. No. 1.5;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```

;
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 493-9300
; TELEFAX: (650) 493-6811
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 500 nucleotides
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
; US-09-324-709A-3
;
Query Match          92.7%; Score 20.4; DB 3; Length 500;
Best Local Similarity 95.5%; Pred. No. 1.4;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GATTGTTGCTGCTGCTCTG 22
Db 80 GATTATTGCTGCTGCTCTG 101

RESULT 4
US-09-018-595B-1
; Sequence 1, Application US/09018595B
; Patent No. 5962233
; GENERAL INFORMATION:
; APPLICANT: Perkin-Elmer Corporation,
; TITLE OF INVENTION: DETERMINATION OF GENOTYPE OF
; TITLE OF INVENTION: AMPLIFICATION PRODUCTS AT MULTIPLE ALLELIC SITES
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David J. Weitz,
; STREET: 650 Page Mill Road
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1050
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch diskette
; OPERATING SYSTEM: Microsoft Windows 95
; SOFTWARE: Wordperfect for windows 6.0,
; SOFTWARE: ASCII (DOS) TEXT format
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/018,595B
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: David J. Weitz
; REGISTRATION NUMBER: 38,362
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 493-9300
; TELEFAX: (650) 493-6811
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 793 nucleotides
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
; US-09-018-595B-1
;
Query Match          92.7%; Score 20.4; DB 2; Length 793;
Best Local Similarity 95.5%; Pred. No. 1.5;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GATTGTTGCTGCTGCTCTG 22
Db 80 GATTATTGCTGCTGCTCTG 101
```

RESULT 5  
US-09-324-709A-1  
; Sequence 1, Application US/09324709A  
; Patent No. 6154707  
; GENERAL INFORMATION:  
; APPLICANT: Perkin-Elmer Corporation,  
; APPLICANT: Applied Biosystems Division  
; TITLE OF INVENTION: FLUORESCENCE GENOTYPING AT MULTIPLE ALLELIC  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: David J. Weitz,  
; ADDRESS: Wilson Sonsini Goodrich & Rosati  
; STREET: 650 Page Mill Road  
; CITY: Palo Alto  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94304-1050  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 inch diskette  
; COMPUTER: IBM compatible  
; OPERATING SYSTEM: Microsoft Windows 95  
; SOFTWARE: Wordperfect for windows 6.0,  
; SOFTWARE: ASCII (DOS) TEXT format  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/324,709A  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: David J. Weitz  
; REGISTRATION NUMBER: 38,362  
; REFERENCE/DOCKET NUMBER: 16842-758  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (650) 493-9300  
; TELEFAX: (650) 493-6811  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 793 nucleotides  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-09-324-709A-1  
Query Match 92.7%; Score 20.4; DB 3; Length 793;  
Best Local Similarity 95.5%; Pred. No. 1.5;  
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 GATTGTGCTGCTGCTCTG 22  
DB 80 GATTGTGCTGCTGCTCTG 101  
RESULT 6  
US-08-899-811-5  
; Sequence 5, Application US/08899811  
; Patent No. 5876942  
; GENERAL INFORMATION:  
; APPLICANT: CHENG, WINSTON T.K.  
; APPLICANT: CHOO, KONG-BUNG  
; APPLICANT: HU, CHE-LIN  
; APPLICANT: WANG, CHIH-HUA  
; APPLICANT: CHEN, CHUAN-WU  
; TITLE OF INVENTION: A PROCESS FOR SEXING COW EMBRYOS  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: BACON & THOMAS  
; STREET: 625 SLATERS LANE - FOURTH FLOOR  
; CITY: ALEXANDRIA

STATE: VIRGINIA  
; COUNTRY: USA  
; ZIP: 22314  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/899,811  
; FILING DATE: 24-JUL-1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: FICHTER, RICHARD E.  
; REGISTRATION NUMBER: 26,382  
; REFERENCE/DOCKET NUMBER: REF/CHEN/881  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 703-683-0500  
; TELEFAX: 703-683-1080  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 20 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; US-08-899-811-5  
Query Match 90.9%; Score 20; DB 2; Length 20;  
Best Local Similarity 100.0%; Pred. No. 1.1;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 ATTGTGCTGCTGCTCTCT 21  
DB 1 ATTGTGCTGCTGCTCTCT 20  
RESULT 7  
US-09-976-594-366  
; Sequence 366, Application US/09976594  
; Patent No. 6673549  
; GENERAL INFORMATION:  
; APPLICANT: Furness, Michael  
; APPLICANT: Buchbinder, Jenny  
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS  
; FILE REFERENCE: PA-0041 US  
; CURRENT APPLICATION NUMBER: US/09/976,594  
; CURRENT FILING DATE: 2001-10-12  
; PRIOR APPLICATION NUMBER: 60/240,409  
; PRIOR FILING DATE: 2000-10-12  
; NUMBER OF SEQ ID NOS: 1143  
; SOFTWARE: PERL Program  
; SEQ ID NO 366  
; LENGTH: 4181  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURES:  
; NAME/KEY: misc. feature  
; OTHER INFORMATION: Incyte ID No. 6673549 337156.2  
; US-09-976-594-366  
Query Match 88.2%; Score 19.4; DB 4; Length 4181;  
Best Local Similarity 95.2%; Pred. No. 6.3;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 2 ATTGTGCTGCTGCTCTCTG 22  
DB 2056 ATTGTGCTGCTGCTCTG 2076  
RESULT 8  
US-08-899-811-4/C  
; Sequence 4, Application US/08899811

```
Patent No. 5876942
GENERAL INFORMATION:
APPLICANT: CHENG, WINSTON T.K.
APPLICANT: CHOO, KONG-BONG
APPLICANT: HU, CHE-LIN
APPLICANT: WANG, CHIH-HUA
APPLICANT: CHEN, CHUAN-MU
TITLE OF INVENTION: A PROCESS FOR SEXING COW EMBRYOS
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: BACON & THOMAS
STREET: 625 SLATERS LANE - FOURTH FLOOR
CITY: ALEXANDRIA
STATE: VIRGINIA
COUNTRY: USA
ZIP: 22314
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/899,811
FILING DATE: 24-JUL-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: FICHTER, RICHARD E.
REGISTRATION NUMBER: 26,382
REFERENCE/DOCKET NUMBER: REF/CHEN/881
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-683-0500
TELEFAX: 703-683-1080
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 19 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-899-811-4

Query Match
Best Local Similarity 86.4%; Score 19; DB 2; Length 19;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 TTTTGTTCCTGCTCCTCT 21
DB 19 TTTTGTTCCTGCTCCTCT 1

RESULT 9
US-09-018-595B-2
Sequence 2, Application US/09018595B
Patent No. 5962233
GENERAL INFORMATION:
APPLICANT: Perkin-Elmer Corporation,
APPLICANT: Applied Biosystems Division
TITLE OF INVENTION: DETERMINATION OF GENOTYPE OF
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: David J. Weltz,
ADDRESSER: Wilson Sonsini Goodrich & Rosati
STREET: 650 Page Mill Road
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1050
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: Microsoft Windows 95
SOFTWARE: Wordperfect for windows 6.0,
SOFTWARE: Wordperfect for windows 6.0,
```

```
SOFTWARE: ASCII (DOS) TEXT format
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/018,595B
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION NUMBER:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: David J. Weltz
REGISTRATION NUMBER: 38,362
REFERENCE/DOCKET NUMBER: PELM-744
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 493-9300
TELEFAX: (650) 493-6811
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 802 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-018-595B-2

Query Match
Best Local Similarity 85.5%; Score 18.8; DB 2; Length 802;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GATTTGTTCCTGCTCCTCTG 22
DB 80 GATTTGTTCCTGCTCCTCTG 101

RESULT 10
US-09-324-709A-2
Sequence 2, Application US/09324709A
Patent No. 6154707
GENERAL INFORMATION:
APPLICANT: Perkin-Elmer Corporation,
APPLICANT: Applied Biosystems Division
TITLE OF INVENTION: FLUORESCENCE GENOTYPING AT MULTIPLE ALLELIC
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: David J. Weltz,
ADDRESSER: Wilson Sonsini Goodrich & Rosati
STREET: 650 Page Mill Road
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1050
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: Microsoft Windows 95
SOFTWARE: Wordperfect for windows 6.0,
SOFTWARE: ASCII (DOS) TEXT format
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/324,709A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION NUMBER:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: David J. Weltz
REGISTRATION NUMBER: 38,362
REFERENCE/DOCKET NUMBER: 16842-758
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 493-9300
TELEFAX: (650) 493-6811
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 802 nucleotides
```

TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-324-709A-2

Query Match 85.5%; Score 18.8; DB 3; Length 802;  
Best Local Similarity 90.9%; Pred. No. 8.5;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GATTTGTTGCTGCTCCTG 22  
DB 80 GATTTGTTGCTGCTCCTG 101

RESULT 11  
US-09-513-999C-35462  
Sequence 35462, Application US/09513999C  
Patent No. 6783961  
GENERAL INFORMATION:  
APPLICANT: Dumas Milne Edwards, J.B.  
APPLICANT: Duclet, A.Y.  
APPLICANT: Giordano, J.Y.  
TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.  
Patent No. 6783961  
FILE REFERENCE: 59,US2,REG  
CURRENT APPLICATION NUMBER: US/09/513,999C  
CURRENT FILING DATE: 2000-02-24  
PRIOR APPLICATION NUMBER: US 60/122,487  
PRIOR FILING DATE: 1999-02-26  
NUMBER OF SEQ ID NOS: 36681  
SOFTWARE: Patent.pm  
SEQ ID NO 35462  
LENGTH: 173  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-513-999C-35462

Query Match 80.9%; Score 17.8; DB 4; Length 173;  
Best Local Similarity 90.5%; Pred. No. 18;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GATTTGTTGCTGCTCCT 21  
DB 13 GATTTGTTGCTGCTCCT 33

RESULT 12  
US-09-338-907-110  
Sequence 110, Application US/09338907  
Patent No. 6265546  
GENERAL INFORMATION:  
APPLICANT: Cohen, Daniel  
APPLICANT: Blumenfeld, Marta  
APPLICANT: Ilya, Chumakov  
APPLICANT: Bougueleret, Lydie  
TITLE OF INVENTION: PROSTATE CANCER GENE  
FILE REFERENCE: GENSET.18CPICP  
CURRENT APPLICATION NUMBER: US/09/338,907  
CURRENT FILING DATE: 1999-06-23  
EARLIER APPLICATION NUMBER: 08/996,306  
EARLIER FILING DATE: 1997-12-22  
EARLIER APPLICATION NUMBER: 60/099,658  
EARLIER FILING DATE: 1998-09-09  
EARLIER APPLICATION NUMBER: 09/218,207  
EARLIER FILING DATE: 1998-12-22  
NUMBER OF SEQ ID NOS: 578  
SOFTWARE: Patent.pm  
SEQ ID NO 110  
LENGTH: 116  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-338-907-110

Query Match 79.1%; Score 17.4; DB 3; Length 116;  
Best Local Similarity 94.7%; Pred. No. 26;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 TTGTTGCTGCTCCTG 22  
DB 72 TTGTTGCTGCTCCTG 90

RESULT 13  
US-09-218-207-110  
Sequence 110, Application US/09218207  
Patent No. 6346381  
GENERAL INFORMATION:  
APPLICANT: Cohen, Daniel  
APPLICANT: Blumenfeld, Marta  
APPLICANT: Ilya, Chumakov  
APPLICANT: Bougueleret, Lydie  
TITLE OF INVENTION: Prostate cancer gene  
FILE REFERENCE: GENSET.018CPI  
CURRENT APPLICATION NUMBER: US/09/218,207  
CURRENT FILING DATE: 1998-12-22  
EARLIER APPLICATION NUMBER: 08/996,306  
EARLIER FILING DATE: 1997-12-22  
EARLIER APPLICATION NUMBER: 60/099,658  
EARLIER FILING DATE: 1998-09-09  
NUMBER OF SEQ ID NOS: 578  
SOFTWARE: Patent.pm  
SEQ ID NO 110  
LENGTH: 116  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-218-207-110

Query Match 79.1%; Score 17.4; DB 3; Length 116;  
Best Local Similarity 94.7%; Pred. No. 26;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 TTGTTGCTGCTCCTG 22  
DB 72 TTGTTGCTGCTCCTG 90

RESULT 14  
US-09-338-907-124  
Sequence 124, Application US/09338907  
Patent No. 6265546  
GENERAL INFORMATION:  
APPLICANT: Cohen, Daniel  
APPLICANT: Blumenfeld, Marta  
APPLICANT: Ilya, Chumakov  
APPLICANT: Bougueleret, Lydie  
TITLE OF INVENTION: PROSTATE CANCER GENE  
FILE REFERENCE: GENSET.18CPICP  
CURRENT APPLICATION NUMBER: US/09/338,907  
CURRENT FILING DATE: 1999-06-23  
EARLIER APPLICATION NUMBER: 08/996,306  
EARLIER FILING DATE: 1997-12-22  
EARLIER APPLICATION NUMBER: 60/099,658  
EARLIER FILING DATE: 1998-09-09  
EARLIER APPLICATION NUMBER: 09/218,207  
EARLIER FILING DATE: 1998-12-22  
NUMBER OF SEQ ID NOS: 578  
SOFTWARE: Patent.pm  
SEQ ID NO 124  
LENGTH: 124  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 31..588  
NAME/KEY: polyA signal  
LOCATION: 5297..5302

OTHER INFORMATION: AATAAA  
US-09-338-907-124

Query Match 79.1%; Score 17.4; DB 3; Length 5326;  
Best Local Similarity 94.7%; Pred. No. 58;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 TTGTTGCTGCTGCTGCTG 22  
|||||  
Db 657 TTGTTGCTGCTGCTGCTG 675

## RESULT 15

US-09-218-207-124  
Sequence 124, Application US/09218207  
Patent No. 6346381  
GENERAL INFORMATION:  
APPLICANT: Cohen, Daniel  
APPLICANT: Blumentfeld, Marta  
APPLICANT: Ilya, Chumakov  
APPLICANT: Bougueleret, Lydie  
TITLE OF INVENTION: Prostate cancer gene  
FILE REFERENCE: GENSET.018CP1  
CURRENT APPLICATION NUMBER: US/09/218.207  
CURRENT FILING DATE: 1998-12-22  
EARLIER APPLICATION NUMBER: 08/996,306  
EARLIER FILING DATE: 1997-12-22  
EARLIER APPLICATION NUMBER: 60/099,658  
EARLIER FILING DATE: 1998-09-09  
NUMBER OF SEQ ID NOS: 578  
SOFTWARE: Patent.pm  
SEQ ID NO 124  
LENGTH: 5324  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 31..588  
NAME/KEY: polyA\_signal  
LOCATION: 5297..5302  
OTHER INFORMATION: AATAAA  
US-09-218-207-124

Query Match 79.1%; Score 17.4; DB 3; Length 5326;  
Best Local Similarity 94.7%; Pred. No. 58;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 TTGTTGCTGCTGCTGCTG 22  
|||||  
Db 657 TTGTTGCTGCTGCTGCTG 675

Search completed: February 8, 2005, 22:25:50  
Job time : 8.0529 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using bw model

Run on: February 8, 2005, 16:37:21, Search time 29.6449 Seconds  
(without alignments)  
4264.119 Million cell updates/sec

Title: US-10-754-437-3

Perfect score: 22  
Sequence: 1 gattctgtctgcctgcctcctg 22

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 4300275 seqs, 2872944193 residues

Total number of hits satisfying chosen parameters: 8600550

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:\*

1: /cgn2\_6/prodata/2/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/prodata/2/pubpna/US06\_NEW\_PUB.seq:\*  
3: /cgn2\_6/prodata/2/pubpna/US06\_PUBCOMB.seq:\*  
4: /cgn2\_6/prodata/2/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/prodata/2/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/prodata/2/pubpna/US07\_PUBCOMB.seq:\*  
7: /cgn2\_6/prodata/2/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/prodata/2/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/prodata/2/pubpna/US09\_PUBCOMB.seq:\*  
10: /cgn2\_6/prodata/2/pubpna/US09\_PUBCOMB.seq:\*  
11: /cgn2\_6/prodata/2/pubpna/US09C\_PUBCOMB.seq:\*  
12: /cgn2\_6/prodata/2/pubpna/US09\_NEW\_PUB.seq:\*  
13: /cgn2\_6/prodata/2/pubpna/US10\_PUBCOMB.seq:\*  
14: /cgn2\_6/prodata/2/pubpna/US10\_PUBCOMB.seq:\*  
15: /cgn2\_6/prodata/2/pubpna/US10C\_PUBCOMB.seq:\*  
16: /cgn2\_6/prodata/2/pubpna/US10\_PUBCOMB.seq:\*  
17: /cgn2\_6/prodata/2/pubpna/US10\_PUBCOMB.seq:\*  
18: /cgn2\_6/prodata/2/pubpna/US10\_NEW\_PUB.seq:\*  
19: /cgn2\_6/prodata/2/pubpna/US10\_NEW\_PUB.seq:\*  
20: /cgn2\_6/prodata/2/pubpna/US60\_NEW\_PUB.seq:\*  
21: /cgn2\_6/prodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the change being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20.4	92.7	65	US-09-908-975-1137	Sequence 1137, App
2	20.4	92.7	500	US-10-104-774-3	Sequence 3, Appl1
3	20.4	92.7	500	US-10-455-150-3	Sequence 3, Appl1
4	20.4	92.7	600	US-10-027-632-287049	Sequence 287049,
5	20.4	92.7	600	US-10-027-632-287050	Sequence 287050,
6	20.4	92.7	600	US-10-027-632-287049	Sequence 287049,
7	20.4	92.7	600	US-10-027-632-287050	Sequence 287050,
8	20.4	92.7	793	US-10-104-774-1	Sequence 1, Appl1
9	20.4	92.7	793	US-10-455-150-1	Sequence 1, Appl1
10	20.4	92.7	793	US-10-755-889-605	Sequence 605, App
11	20.4	92.7	259202	US-10-723-860-1311	Sequence 1311, App
12	19.4	88.2	228	US-10-674-124A-17105	Sequence 17105, A

13	19.4	88.2	2190	US-10-295-027-581	Sequence 581, App
14	19.4	88.2	2681	US-10-789-090-1	Sequence 1, Appl1
15	19.4	88.2	2685	US-09-964-824A-257	Sequence 257, App
16	19.4	88.2	2685	US-10-205-823-366	Sequence 366, App
17	19.4	88.2	2685	US-10-757-262-19	Sequence 19, Appl1
18	18.8	85.5	802	US-10-104-774-2	Sequence 2, Appl1
19	18.8	85.5	802	US-10-455-150-2	Sequence 2, Appl1
20	18.8	85.5	852	US-10-723-860-5797	Sequence 5797, App
21	18.4	83.6	1287	US-09-764-847-144	Sequence 144, App
22	18.4	83.6	1287	US-10-092-154-144	Sequence 144, App
23	18.4	83.6	1295	US-09-764-847-495	Sequence 495, App
24	18.4	83.6	1295	US-10-092-154-495	Sequence 495, App
25	18.4	83.6	168276	US-10-087-192-142	Sequence 142, App
26	18	81.8	668	US-10-027-632-34029	Sequence 34029, A
27	18	81.8	668	US-10-027-632-34029	Sequence 34029, A
28	17.8	80.9	598	US-10-027-632-217666	Sequence 217666,
29	17.8	80.9	598	US-10-027-632-217667	Sequence 217667,
30	17.8	80.9	598	US-10-027-632-217667	Sequence 217667,
31	17.8	80.9	598	US-10-027-632-217666	Sequence 217666,
32	17.8	80.9	598	US-10-027-632-217667	Sequence 217667,
33	17.8	80.9	598	US-10-027-632-217667	Sequence 217667,
34	17.4	79.1	80	US-10-384-245-267	Sequence 267, App
35	17.4	79.1	116	US-09-901-484A-110	Sequence 110, App
36	17.4	79.1	116	US-09-901-484A-110	Sequence 110, App
37	17.4	79.1	349	US-10-674-124A-26540	Sequence 26540, A
38	17.4	79.1	2756	US-10-062-574-1625	Sequence 1625, App
39	17.4	79.1	3923	US-10-152-319A-1931	Sequence 1931, App
40	17.4	79.1	5324	US-09-901-484A-124	Sequence 124, App
41	17.4	79.1	5324	US-09-853-526-124	Sequence 124, App
42	17.4	79.1	56516	US-09-901-484A-1	Sequence 1, Appl1
43	17.4	79.1	56520	US-09-853-526-1	Sequence 1, Appl1
44	17.4	79.1	56520	US-09-901-484A-179	Sequence 179, App
45	17.4	79.1	56520	US-09-853-526-179	Sequence 179, App

#### ALIGNMENTS

RESULT 1

US-09-908-975-1137

Sequence 1137, Application US/0908975

Publication No. US20030165843A1

GENERAL INFORMATION:

APPLICANT: SHOSHAN, Avi

APPLICANT: WASSERMAN, Alon

APPLICANT: MINTZ, Eli

APPLICANT: FAIGER, Shimon

TITLE OF INVENTION: OLIGONUCLEOTIDE LIBRARY FOR DETECTING RNA TRANSCRIPTS AND SPLIC

TITLE OF INVENTION: THAT POPULATE A TRANSCRIPTOME

CURRENT APPLICATION NUMBER: US/09/908,975

CURRENT FILING DATE: 2001-07-20

PRIOR APPLICATION NUMBER: US 60/287,724

PRIOR FILING DATE: 2001-05-02

PRIOR APPLICATION NUMBER: US 60/221,607

PRIOR FILING DATE: 2000-07-28

NUMBER OF SEQ ID NOS: 32337

SOFTWARE: PatentIn version 3.0

SEQ ID NO 1137

LENGTH: 65

TYPE: DNA

ORGANISM: Rattus norvegicus

US-09-908-975-1137

Query Match 92.7% Score 20.4; DB 10; Length 65;  
Best Local Similarity 95.5%; Pred. No. 9; 6;  
Matches 21; Conservative 0; Indels 0; Gaps 0;

1 GATTGTGTTGCTGCTGCTGCTG 22  
|||||  
6 GATCTGTGCTGCTGCTGCTGCTG 27

```
RESULT 2
US-10-104-774-3
; Sequence 3, Application US/10104774
; Publication No. US20020164630A1
; GENERAL INFORMATION:
; APPLICANT: Perkin-Elmer Corporation, Applied Biosystems Division
; TITLE OF INVENTION: DETERMINATION OF GENOTYPE OF AMPLIFICATION PRODUCTS AT MULTIPLE A
; FILE REFERENCE: 16842-782
; CURRENT APPLICATION NUMBER: US/10/104,774
; CURRENT FILING DATE: 2002-03-21
; PRIOR APPLICATION NUMBER: US 09/018,595
; PRIOR FILING DATE: 1998-02-04
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 500
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-104-774-3

Query Match          92.7%; Score 20.4; DB 13; Length 500;
Best Local Similarity 95.5%; Pred. No. 11;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 GATTGTGTCCTGCTCCTG 22
        ||||| ||||| ||||| |||||
Db      80 GATTATTGCTGCTCCTG 101

RESULT 3
US-10-455-150-3
; Sequence 3, Application US/10455150
; Publication No. US2004005302A1
; GENERAL INFORMATION:
; APPLICANT: Perkin-Elmer Corporation,
; TITLE OF INVENTION: FLUORESCENCE GENOTYPING AT MULTIPLE ALLELIC
; SITES
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David J. Weltz,
; STREET: 650 Page Mill Road
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1050
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch diskette
; OPERATING SYSTEM: Microsoft Windows 95
; SOFTWARE: Wordperfect for windows 6.0,
; ASCII (DOS) TEXT format
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/455,150
; FILING DATE: 04-Jun-2003
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/324,709A
; FILING DATE: 03-June-1999
; ATTORNEY/AGENT INFORMATION:
; NAME: David J. Weltz
; REGISTRATION NUMBER: 38,362
; REFERENCE/DOCKET NUMBER: 16842-758
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 493-9300
; TELEFAX: (650) 493-6811
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 500 nucleotides
; TYPE: nucleic acid
```

```
STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-10-455-150-3

Query Match          92.7%; Score 20.4; DB 16; Length 500;
Best Local Similarity 95.5%; Pred. No. 11;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 GATTGTGTCCTGCTCCTG 22
        ||||| ||||| ||||| |||||
Db      245 GATTATTGCTGCTCCTG 224

RESULT 5
US-10-027-632-287049/c
; Sequence 287049, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 287049
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-287049

Query Match          92.7%; Score 20.4; DB 13; Length 600;
Best Local Similarity 95.5%; Pred. No. 11;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 GATTGTGTCCTGCTCCTG 22
        ||||| ||||| ||||| |||||
Db      245 GATTATTGCTGCTCCTG 224
```



```

; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 287050
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-287050
```

```
Query Match          92.7%; Score 20.4; DB 13; Length 600;
Best Local Similarity 95.5%; Pred. No. 11;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 GATTTGTTTGCTGCTGCTG 22
DB 245 GATTTATTGCTGCTGCTG 224
```

```
RESULT 6
US-10-027-632-287049/c
; Sequence 287049, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
```

```

; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 287049
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-287049
```

```
Query Match          92.7%; Score 20.4; DB 15; Length 600;
Best Local Similarity 95.5%; Pred. No. 11;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 GATTTGTTTGCTGCTGCTG 22
DB 245 GATTTATTGCTGCTGCTG 224
```

```
RESULT 7
US-10-027-632-287050/c
; Sequence 287050, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
```

```

; APPLICANT: Wang, David G.
```

```

; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 287050
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-287050
```

```
Query Match          92.7%; Score 20.4; DB 15; Length 600;
Best Local Similarity 95.5%; Pred. No. 11;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 GATTTGTTTGCTGCTGCTG 22
DB 245 GATTTATTGCTGCTGCTG 224
```

```
RESULT 8
US-10-104-774-1
; Sequence 1, Application US/10104774
; Publication No. US20020164630A1
; GENERAL INFORMATION:
```

```

; APPLICANT: Perkin-Elmer Corporation, Applied Biosystems Division
; TITLE OF INVENTION: DETERMINATION OF GENOTYPE OF AMPLIFICATION PRODUCTS AT MULTIPLE
; FILE REFERENCE: 16842-782
; CURRENT APPLICATION NUMBER: US/10/104,774
; PRIOR FILING DATE: 2002-03-21
; PRIOR APPLICATION NUMBER: US 09/018,595
; PRIOR FILING DATE: 1998-02-04
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 793
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-104-774-1
```

```
Query Match          92.7%; Score 20.4; DB 13; Length 793;
Best Local Similarity 95.5%; Pred. No. 11;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 GATTTGTTTGCTGCTGCTG 22
DB 80 GATTTATTGCTGCTGCTG 101
```

```
RESULT 9
US-10-455-150-1
; Sequence 1, Application US/10455150
; Publication No. US20040053302A1
; GENERAL INFORMATION:
```

```

; APPLICANT: Perkin-Elmer Corporation,
; Applied Biosystems Division
```

TITLE OF INVENTION: FLUORESCENCE GENOTYPING AT MULTIPLE ALLELIC  
SITES  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David J. Weitz,  
Wilson Sonsini Goodrich & Rosati  
STREET: 650 Page Mill Road  
CITY: Palo Alto  
STATE: California  
COUNTRY: USA  
ZIP: 94304-1050  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch diskette  
COMPUTER: IBM compatible  
OPERATING SYSTEM: Microsoft Windows 95  
SOFTWARE: Wordperfect for windows 6.0,  
ASCII (DOS) TEXT format  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/455,150  
FILING DATE: 04-Jun-2003  
CLASSIFICATION: <unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/324,709A  
FILING DATE: 03-June-1999  
ATTORNEY/AGENT INFORMATION:  
NAME: David J. Weitz  
REGISTRATION NUMBER: 38,362  
REFERENCE/DOCKET NUMBER: 16842-758  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 493-9300  
TELEFAX: (650) 493-6811  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 793 nucleotides  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-10-455-150-1

Query Match 92.7%; Score 20.4; DB 16; Length 793;  
Best Local Similarity 95.5%; Pred. No. 11;  
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GATTTGTTGGCTGCTCCTG 22  
|||||  
DB 80 GATTTATTGGCTGCTCCTG 101

RESULT 10  
US-10-755-889-605  
Sequence 605, Application US/10755889  
Publication No. US20040171823A1  
GENERAL INFORMATION:  
APPLICANT: Bristol-Myers Squibb Company  
TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ASSOCIATED WITH THE NF-KB  
FILE REFERENCE: D0284 NP  
CURRENT APPLICATION NUMBER: US/10/755,889  
PRIOR FILING DATE: 2004-01-13  
PRIOR APPLICATION NUMBER: U.S. 60/440,068  
PRIOR FILING DATE: 2003-01-14  
PRIOR APPLICATION NUMBER: U.S. 60/469,757  
PRIOR FILING DATE: 2003-05-12  
NUMBER OF SEQ ID NOS: 823  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 605  
LENGTH: 793  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-755-889-605

Query Match 92.7%; Score 20.4; DB 17; Length 793;

Best Local Similarity 95.5%; Pred. No. 11;  
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GATTTGTTGGCTGCTCCTG 22  
|||||  
DB 80 GATTTATTGGCTGCTCCTG 101

RESULT 11  
US-10-723-860-1311  
Sequence 1311, Application US/10723860  
Publication No. US20040253606A1  
GENERAL INFORMATION:  
APPLICANT: Aziz, Nacasha  
APPLICANT: Ginsburg, Wendy M.  
APPLICANT: Zlotnik, Albert  
TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &  
FILE REFERENCE: 05892,0193,NPUS01  
CURRENT APPLICATION NUMBER: US/10/723,860  
PRIOR FILING DATE: 2003-11-26  
PRIOR APPLICATION NUMBER: 60/429,739  
PRIOR FILING DATE: 2002-11-26  
NUMBER OF SEQ ID NOS: 8393  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 1311  
LENGTH: 259202  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-723-860-1311

Query Match 92.7%; Score 20.4; DB 18; Length 259202;  
Best Local Similarity 95.5%; Pred. No. 15;  
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GATTTGTTGGCTGCTCCTG 22  
|||||  
DB 69884 GATTTATTGGCTGCTCCTG 69905

RESULT 12  
US-10-674-124A-17105  
Sequence 17105, Application US/10674124A  
Publication No. US20040197979A1  
GENERAL INFORMATION:  
APPLICANT: INOKO, Hidetoshi  
APPLICANT: TANIYA, Gen  
TITLE OF INVENTION: GENE MAPPING METHOD USING MICROSATELLITE  
FILE REFERENCE: ORIN-003CIP  
CURRENT APPLICATION NUMBER: US/10/674,124A  
PRIOR FILING DATE: 2003-09-26  
PRIOR APPLICATION NUMBER: 10/257,511  
PRIOR FILING DATE: 2003-03-07  
PRIOR APPLICATION NUMBER: PCT/JP00/07621  
PRIOR FILING DATE: 2000-10-30  
PRIOR APPLICATION NUMBER: JP2000-112639  
PRIOR FILING DATE: 2000-04-13  
PRIOR APPLICATION NUMBER: JP2002-327516  
PRIOR FILING DATE: 2002-09-28  
PRIOR APPLICATION NUMBER: JP2002-383869  
PRIOR FILING DATE: 2002-12-09  
NUMBER OF SEQ ID NOS: 27110  
SEQ ID NO 17105  
LENGTH: 228  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
FEATURE: 267364  
OTHER INFORMATION: 267364  
OTHER INFORMATION: Located on chromosome 11  
FEATURE:  
OTHER INFORMATION: Distance between a terminus base of telomere on

OTHER INFORMATION: chromosomal short arm and 5'-terminus of this base  
OTHER INFORMATION: sequence : 13242364  
FEATURE:  
OTHER INFORMATION: Distance between 3'-terminus of neighbour sequence of  
OTHER INFORMATION: sequence listing upward to telomere on chromosomal short arm and  
OTHER INFORMATION: 5'-terminus of this base sequence : 225419  
US-10-674-124A-17105

Query Match 88.2%; Score 19.4; DB 18; Length 228;  
Best Local Similarity 95.2%; Pred. No. 30;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
2 ATTTGTTGGCTGCTCTCG 22  
191 ATTTTGTTCCTGCTCTCG 211

RESULT 13  
US-10-295-027-581  
Sequence 581, Application US/10795027  
Publication No. US2003023350A1  
GENERAL INFORMATION:  
APPLICANT: Afar, Daniel  
APPLICANT: Aziz, Natascha  
APPLICANT: Gansberg, Wendy M.  
APPLICANT: Gish, Kurt C.  
APPLICANT: Glynn, Richard  
APPLICANT: Hevezi, Peter A.  
APPLICANT: Mack, David H.  
APPLICANT: Murray, Richard  
APPLICANT: Watson, Susan R.  
APPLICANT: Eos Biotechnology, Inc.  
TITLE OF INVENTION: Methods of Diagnosis of Cancer: Compositions and  
TITLE OF INVENTION: Methods of Screening for Modulators of Cancer  
FILE REFERENCE: 018501-012500US  
CURRENT FILING DATE: 2002-11-13  
PRIOR APPLICATION NUMBER: US/10/295,027  
PRIOR FILING DATE: 2000-09-15  
PRIOR APPLICATION NUMBER: US 09/663,733  
PRIOR FILING DATE: 2000-09-15  
PRIOR APPLICATION NUMBER: US 60/350,666  
PRIOR FILING DATE: 2001-11-13  
PRIOR APPLICATION NUMBER: US 60/335,394  
PRIOR FILING DATE: 2001-11-15  
PRIOR APPLICATION NUMBER: US 60/332,464  
PRIOR FILING DATE: 2001-11-21  
PRIOR APPLICATION NUMBER: US 60/334,393  
PRIOR FILING DATE: 2001-11-29  
PRIOR APPLICATION NUMBER: US 60/340,376  
PRIOR FILING DATE: 2001-12-14  
PRIOR APPLICATION NUMBER: US 60/347,211  
PRIOR FILING DATE: 2002-01-08  
PRIOR APPLICATION NUMBER: US 60/347,349  
PRIOR FILING DATE: 2002-01-10  
PRIOR APPLICATION NUMBER: US 60/355,250  
PRIOR FILING DATE: 2002-02-08  
PRIOR APPLICATION NUMBER: US 60/356,714  
PRIOR FILING DATE: 2002-02-13  
Remaining prior Application data removed - See file Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 1386  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 581  
LENGTH: 2190  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-295-027-581

Query Match 88.2%; Score 19.4; DB 15; Length 2190;  
Best Local Similarity 95.2%; Pred. No. 34;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
2 ATTTGTTGGCTGCTCTCG 22  
206 ATTTTGTTCCTGCTCTCG 2046

RESULT 14  
US-10-789-090-1  
Sequence 1, Application US/10789090  
Publication No. US20040223970A1  
GENERAL INFORMATION:  
APPLICANT: Afar, Daniel  
APPLICANT: Law, Debbie  
TITLE OF INVENTION: ANTIBODIES AGAINST SLC15A2 AND USES THEREOF  
FILE REFERENCE: 05882.0192.NPUS01  
CURRENT APPLICATION NUMBER: US/10/789,090  
CURRENT FILING DATE: 2004-02-27  
NUMBER OF SEQ ID NOS: 10  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 1  
LENGTH: 2681  
TYPE: DNA  
ORGANISM: Homo Sapiens  
US-10-789-090-1

Query Match 88.2%; Score 19.4; DB 18; Length 2681;  
Best Local Similarity 95.2%; Pred. No. 35;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

2 ATTTGTTGGCTGCTCTCG 22  
2055 ATTTGTTTCTGCTCTCG 2075

RESULT 15  
US-09-964-824A-257  
Sequence 257, Application US/09964824A  
Patent No. US20020102531A1  
GENERAL INFORMATION:  
APPLICANT: Horrigan, Stephen  
TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signat  
TITLE OF INVENTION: Sets  
FILE REFERENCE: 689290-73  
CURRENT APPLICATION NUMBER: US/09/964,824A  
CURRENT FILING DATE: 2001-09-27  
PRIOR APPLICATION NUMBER: US/60/236,033  
PRIOR FILING DATE: 2000-09-28  
PRIOR APPLICATION NUMBER: US/60/236,032  
PRIOR FILING DATE: 2000-09-28  
PRIOR APPLICATION NUMBER: US/60/236,028  
PRIOR FILING DATE: 2000-09-28  
NUMBER OF SEQ ID NOS: 583  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 257  
LENGTH: 2685  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-964-824A-257

Query Match 88.2%; Score 19.4; DB 9; Length 2685;  
Best Local Similarity 95.2%; Pred. No. 35;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
2 ATTTGTTGGCTGCTCTCG 22  
2056 ATTTGTTTCTGCTCTCG 2076

Search completed: February 9, 2005, 06:50:52  
Job time : 31.6449 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 8, 2005, 12:58:25 ; Search time 262.041 Seconds  
(without alignments)  
3059.349 Million cell updates/sec

Title: US-10-754-437-3

Perfect score: 22  
Sequence: 1 gattctgtctgtctgtctctctg 22

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*  
1: gb\_esc1.\*  
2: gb\_esc2.\*  
3: gb\_hnc.\*  
4: gb\_esc3.\*  
5: gb\_esc4.\*  
6: gb\_esc5.\*  
7: gb\_esc6.\*  
8: gb\_g881.\*  
9: gb\_g882.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	22	100.0	303	7	W33932
2	22	100.0	430	7	W40649
3	22	100.0	536	7	W36345
4	22	100.0	585	9	AY419441
5	22	100.0	645	6	CD773419
6	22	100.0	707	6	CB056709
7	22	100.0	812	3	AK029358
8	22	100.0	843	6	CB588212
9	22	100.0	859	6	CB588525
10	22	100.0	865	6	CB589177
11	22	100.0	886	6	CB587051
12	22	100.0	889	6	CB589251
13	22	100.0	891	6	CB574837
14	22	100.0	913	6	CB590451
15	22	100.0	919	6	CB587332
16	22	100.0	928	6	CB590111
17	22	100.0	945	2	BB614068
18	22	100.0	990	7	W12906
19	22	100.0	1006	7	W29475
20	22	100.0	1020	7	W08102
21	20.4	92.7	217	7	R46913
22	20.4	92.7	223	7	BB571643
23	20.4	92.7	257	7	R47100
24	20.4	92.7	429	6	CB473334

25	20.4	92.7	449	7	R47024
26	20.4	92.7	556	7	R46934
27	20.4	92.7	570	9	AY419439
28	20.4	92.7	570	9	AY419440
29	20.4	92.7	1031	6	CB575508
30	19.4	88.2	286	7	R46887
31	19.4	88.2	293	7	R47143
32	19.4	88.2	298	7	R45420
33	19.4	88.2	303	7	R46903
34	19.4	88.2	318	7	R47078
35	19.4	88.2	355	7	R46940
36	19.4	88.2	355	7	R46941
37	19.4	88.2	373	7	R47135
38	19.4	88.2	395	7	R47030
39	19.4	88.2	458	2	AM295942
40	19.4	88.2	461	1	AA054054
41	19.4	88.2	467	7	R46947
42	19.4	88.2	523	1	AA649247
43	19.4	88.2	621	2	AM974548
44	19.4	88.2	656	5	BQ017433
45	19.4	88.2	699	7	CF366828

## ALIGNMENTS

RESULT 1  
LOCUS W33932  
DEFINITION mb54602.r1 Soares mouse p3NMF19.5 Mus musculus cDNA clone IMAGE:333219 5' similar to gb:M10095 (Mouse amelogenin (MOUSE)).  
RNA sequence.  
W33932  
W33932.1 GI:1315837  
EST.  
Mus musculus (house mouse)  
Mus musculus  
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 303)  
Marras, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Stepien, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.  
The WashU-HMI Mouse EST Project  
Unpublished (1996)  
Contact: Maria M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.edu  
This clone is available royalty-free through INM; contact the IMAGE Consortium (info@image.lim.gov) for further information.  
WGI:214619  
Seq primer: mob. RBGA+ET  
High quality sequence stop: 185.  
Location/Qualifiers  
1..303  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:333219"  
/dev stage="19.5 dpc total fetus"  
/lab host="DH10B (ampicillin resistant)"  
/clone lib="Soares mouse p3NMF19.5"  
/note="Vector: pRT73D (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGAGCGCGCATTTTCTTTTCTTTT 3'] double-stranded cDNA was size selected, ligated to Eco RI

## ORIGIN

adapters (pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pRT3 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M. Fatima Bonaldo. RNA was kindly provided by Dr. Minoru Ko (Wayne State University)."

Query Match 100.0%; Score 22; DB 7; Length 303;  
Best Local Similarity 100.0%; Pred. No. 29;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Cy 1 GATTGTGTCCTGCTCTG 22  
|||||  
Db 81 GATTGTGTCCTGCTCTG 102

RESULT 2 430 bp mRNA linear EST 11-SEP-1996  
W40649  
LOCUS  
DEFINITION  
IMAGE:351335.5 similar to gb:M10095 Mouse amelogenin (MOUSE);  
mRNA sequence.  
W40649.1 GI:1324982  
EST.  
Mus musculus (house mouse)  
ORGANISM  
Mus musculus (house mouse)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 430)  
Marras, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
Waterston, R.  
The WashU-HMI Mouse EST Project  
Unpublished (1996)  
Contact: Marra M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watsn.wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:223135  
Seq primer: ETPRimer  
High quality sequence stop: 323.  
Location/Qualifiers  
1. 430  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:351335"  
/dev\_stage="19.5 dpc total fetus"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Vector: pRT3D (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
was primed with a Not I - oligo(dT) primer [5'  
TGTTACCAATCTGAAGTGGAGCGCGCATTTTCTTTTCTTTT 3']  
double-stranded cDNA was size selected, ligated to Eco RI  
adapters (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of a modified pRT3 vector  
(Pharmacia). Library went through one round of  
normalization to a Cot = 5. Library constructed by Bento  
Soares and M. Fatima Bonaldo. RNA was kindly provided by  
Dr. Minoru Ko (Wayne State University)."

## FEATURES

source

TITLE  
JOURNAL  
COMMENT

## REFERENCE

AUTHORS

Query Match 100.0%; Score 22; DB 7; Length 303;  
Best Local Similarity 100.0%; Pred. No. 29;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Cy 1 GATTGTGTCCTGCTCTG 22  
|||||  
Db 81 GATTGTGTCCTGCTCTG 102

RESULT 2 430 bp mRNA linear EST 11-SEP-1996  
W40649  
LOCUS  
DEFINITION  
IMAGE:351335.5 similar to gb:M10095 Mouse amelogenin (MOUSE);  
mRNA sequence.  
W40649.1 GI:1324982  
EST.  
Mus musculus (house mouse)  
ORGANISM  
Mus musculus (house mouse)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 430)  
Marras, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
Waterston, R.  
The WashU-HMI Mouse EST Project  
Unpublished (1996)  
Contact: Marra M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watsn.wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:223135  
Seq primer: ETPRimer  
High quality sequence stop: 323.  
Location/Qualifiers  
1. 430  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:351335"  
/dev\_stage="19.5 dpc total fetus"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Vector: pRT3D (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
was primed with a Not I - oligo(dT) primer [5'  
TGTTACCAATCTGAAGTGGAGCGCGCATTTTCTTTTCTTTT 3']  
double-stranded cDNA was size selected, ligated to Eco RI  
adapters (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of a modified pRT3 vector  
(Pharmacia). Library went through one round of  
normalization to a Cot = 5. Library constructed by Bento  
Soares and M. Fatima Bonaldo. RNA was kindly provided by  
Dr. Minoru Ko (Wayne State University)."

## ORIGIN

adapters (pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pRT3 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M. Fatima Bonaldo. RNA was kindly provided by Dr. Minoru Ko (Wayne State University)."

Query Match

100.0%; Score 22; DB 7; Length 430;

Best Local Similarity 100.0%; Pred. No. 30;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Cy 1 GATTGTGTCCTGCTCTG 22  
|||||  
Db 69 GATTGTGTCCTGCTCTG 90

RESULT 3 536 bp mRNA linear EST 11-SEP-1996  
W36345  
LOCUS  
DEFINITION  
IMAGE:334966.5 similar to gb:M10095 Mouse amelogenin (MOUSE);  
mRNA sequence.  
W36345.1 GI:1318120  
EST.  
Mus musculus (house mouse)  
ORGANISM  
Mus musculus (house mouse)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 536)  
Marras, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
Waterston, R.  
The WashU-HMI Mouse EST Project  
Unpublished (1996)  
Contact: Marra M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watsn.wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:216366  
Seq primer: ETPRimer  
High quality sequence stop: 359.  
Location/Qualifiers  
1. 536  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:334966"  
/dev\_stage="19.5 dpc total fetus"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Vector: pRT3D (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
was primed with a Not I - oligo(dT) primer [5'  
TGTTACCAATCTGAAGTGGAGCGCGCATTTTCTTTTCTTTT 3']  
double-stranded cDNA was size selected, ligated to Eco RI  
adapters (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of a modified pRT3 vector  
(Pharmacia). Library went through one round of  
normalization to a Cot = 5. Library constructed by Bento  
Soares and M. Fatima Bonaldo. RNA was kindly provided by  
Dr. Minoru Ko (Wayne State University)."

## FEATURES

source

## ORIGIN

source

Query Match 100.0%; Score 22; DB 7; Length 536;  
Best Local Similarity 100.0%; Pred. No. 31;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Cy 1 GATTGTGTCCTGCTCTG 22  
|||||  
Db 64 GATTGTGTCCTGCTCTG 85

RESULT 4

LOCUS	AY19441	585 bp	DNA	linear	GSS 17-DEC-2003
DEFINITION	Mus musculus AMELX gene, VIRTUAL TRANSCRIPT, partial sequence,				
ACCESSION	AY19441				
VERSION	AY19441.1	GI:3975398			
KEYWORDS	GSS.				
SOURCE	Mus musculus (house mouse)				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
AUTHORS	1 (bases 1 to 585)				
	Todd,M.A., Tanenbaum,D.M., Civeello,D.R., Lu,F., Murphy,B., Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Slnsky,J.J., Adams,M.D. and Cargill,M.				
	Interfering nonneutral evolution from human-chimp-mouse orthologous gene trios				
TITLE	Science 302 (5652), 1960-1963 (2003)				
JOURNAL	Science 302 (5652), 1960-1963 (2003)				
PUBMED	14671302				
REFERENCE	2 (bases 1 to 585)				
AUTHORS	Clark,A.G., Gnanowski,S., Nielson,R., Thomas,P., Kejarival,A., Todd,M.A., Tanenbaum,D.M., Civeello,D.R., Lu,F., Murphy,B., Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Slnsky,J.J., Adams,M.D. and Cargill,M.				
	Direct Submission				
	Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA				
COMMENT	These sequences were made by sequencing genomic exons and ordering them based on alignment.				
FEATURES	Location/Qualifiers				
source	1..585				
	/organism="Mus musculus"				
	/mol_type="genomic DNA"				
	/db_xref="taxon:10090"				
	<1..>585				
	/gene="AMELX"				
	/locus_tag="HCM6891"				
ORIGIN					
	Query Match	100.0%;	Score 22;	DB 9;	Length 585;
	Best local similarity	100.0%;	Pred. No. 31;		
	Matches 22;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1 GATTGTGTTGGCTGCTCTCTG 22				
	12 GATTGTGTTGGCTGCTCTCTG 33				
DB					
RESULT 5					
CD773419	645 bp	mRNA	linear	EST 02-JUL-2003	
DEFINITION	AGNSCOURT 14713373 NIH MGC 190 Mus musculus cDNA clone				
LOCUS	IMAGE:30501267 5', mRNA sequence.				
ACCESSION	CD773419				
VERSION	CD773419.1	GI:32431921			
KEYWORDS	EST.				
SOURCE	Mus musculus (house mouse)				
ORGANISM	Mus musculus				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
	1 (bases 1 to 645)				
	NIH-MGC http://mgc.nci.nih.gov/.				
	National Institutes of Health, Mammalian Gene Collection (MGC)				
	Unpublished (1999)				
REFERENCE	Contact: Daniela S. Gerhard, Ph.D.				
AUTHORS	Office of Cancer Genomics				
TITLE	National Cancer Institute / NIH				
	Bldg. 31 Rm10A07 Bethesda, MD 20892				
	Email: cga@bcr-remail.nih.gov				
	Tissue Procurement: Yoshiniko Yamada, Takeehi Nakamura, NIDCR				
	cDNA library Preparation: CLOWTECH Laboratories, Inc.				
COMMENT	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLNL)				

DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LINL at:  
<http://image.llnl.gov>  
 Plate: NDCM223 row: k column: 04  
 High quality sequence at: 611.  
 Location/Qualifiers  
 1..645  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:30501267"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_id="NH\_MGC\_190"  
 /note="Organ: Pooled - Molar; Vector: pDNR-LIB; Site: 1:  
 SfiI (ggccgctcgcc); Site 2: SfiI (ggccgctcgcc);  
 Non-normalized full-length enriched library 5' and 3'  
 adaptors were used in cloning as follows: 5' adaptor  
 sequence: 5'-CACGCGCATTTTGGCC-3' and 3' adaptor sequence:  
 5'-ATTCTAGAGCGCCGAGCGCGCCGACATG-dT(30)BN-3' (where B = A,  
 C, or G and N = A, C, G, or T). Average insert size 1.71  
 kb (range 0.5-3.0 kb). 15/15 colonies contained inserts by  
 PCR. This library was enriched for full-length clones and  
 was constructed by Clontech Laboratories (Palo Alto, CA)  
 Corp."

ORIGIN  
 Query Match 100.0%; Score 22; DB 6; Length 645;  
 Best Local Similarity 100.0%; Pred. No. 31;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 GATTTGTTGGCTGCTCCTCG 22  
 |||||||||  
 Db 84 GATTTGTTGGCTGCTCCTCG 105

RESULT 6  
 CB056709  
 LOCUS  
 DEFINITION  
 707 bp mRNA linear EST 17-JAN-2003  
 NISC j119g05.w1 Soares NMBP13-15 Mus musculus cDNA clone  
 IMAGE:4848584 5', mRNA sequence.  
 CB056709  
 CB056709.1 GI:27794996  
 EST.

ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Mus musculus (house mouse)  
 Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 707)  
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished (1997)  
 Contact: Robert Straubeberg, Ph.D.  
 Email: cgaps-rcmail.nih.gov  
 CDNA Library Preparation:  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium/LINL  
 DNA Sequencing by: National Institutes of Health Intramural  
 Sequencing Center (NISC)  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LINL at:  
<http://image.llnl.gov>  
 Plate: LHAM10779 row: N column: 9  
 Seq primer: T7 primer.

JOURNAL  
 COMMENT  
 Location/Qualifiers  
 1..707  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:4848584"  
 /issue\_type="pituitary gland"  
 /dev\_stage="juvenile, 13-15 days"  
 /lab\_host="DH10B (phage-resistant)"

FEATURES  
 SOURCE

/clone.lib="Soares NMBP13-15"  
 /note="Organ: brain; Vector: pT73D-Pac1; Site 1: NotI;  
 Site 2: EcoRI; 1st strand cDNA was primed with a NotI-  
 oligo (DT) primer  
 5'-ACTGGAAGATTCGCGCGCCGCTGACCGATGTTTCTTTTCTTTT-3';  
 double-stranded cDNA was ligated to EcoRI adaptors  
 5'-ATTGCGCAGCAG-3' and 5'-CCTGCGCG-3' (Pharmacia),  
 digested with NotI and cloned into the NotI and EcoRI  
 sites of the pT73D-Pac1 vector. Library went through one  
 round of normalization, and was constructed in the  
 laboratory of M. Bento Soares (University of Iowa)."

## ORIGIN

Query Match 100.0%; Score 22; DB 6; Length 707;  
 Best Local Similarity 100.0%; Pred. No. 32;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GATTGTGTTGCTGCTGCTGCTG 22  
 58 GATTGTGTTGCTGCTGCTGCTG 79

## Db

RESULT 7  
 AK029358 812 bp mRNA linear HTC 03-APR-2004  
 LOCUS AK029358  
 DEFINITION Mus musculus 0 day neonate head cDNA, RIKEN full-length enriched  
 library, clone:4833404E21 product:amelogenin, full insert sequence.  
 ACCESSION AK029358  
 VERSION AK029358.1 GI:26325321  
 KEYWORDS HTC; CAP trapper.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
 AUTHORS Carninci, P. and Hayashizaki, Y.  
 TITLE High-efficiency full-length cDNA cloning  
 JOURNAL Meth. Enzymol. 303, 19-44 (1999)  
 MEDLINE 99279253  
 PUBMED 10349636

## REFERENCE

AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
 Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
 TITLE Normalization and subtraction of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new genes  
 JOURNAL Genome Res. 10 (10), 1617-1630 (2000)  
 MEDLINE 20499374  
 PUBMED 11042159

REFERENCE  
 AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,  
 Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M.,  
 Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishise, T., Harada, A.,  
 Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,  
 Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matshiki, M.,  
 Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,  
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.  
 TITLE RIKEN integrated sequence analysis (RISA) system--384-format  
 JOURNAL sequencing pipeline with 384 multicapillary sequencer  
 MEDLINE 20530913  
 PUBMED 11076861

REFERENCE  
 AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the  
 FANTOM Consortium.  
 TITLE Functional annotation of a full-length mouse cDNA collection  
 JOURNAL Nature 409, 685-690 (2001)  
 REFERENCE 5

AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research  
 Group Phase I & II Team.  
 TITLE Analysis of the mouse transcriptome based on functional annotation  
 JOURNAL of 60,770 full-length cDNAs  
 MEDLINE Nature 420, 563-573 (2002)  
 REFERENCE 6 (bases 1 to 812)

## AUTHORS

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,  
 Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,  
 Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,  
 Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,  
 Katoh, H., Kawai, J., Kojima, K., Kondo, S., Konno, H., Konda, M.,  
 Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,  
 Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohnato, N.,  
 Okazaki, Y., Saito, R., Saitoh, K., Sakai, C., Sakai, K., Sakazume, N.,  
 Sano, H., Sasaki, D., Shibata, K., Shingawa, A., Shiraki, T.,  
 Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,  
 Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,  
 Muramatsu, M. and Hayashizaki, Y.

## TITLE

Submitted (16-JUN-2001) Yoshihide Hayashizaki, The Institute of  
 Physical and Chemical Research (RIKEN), Laboratory for Genome  
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
 RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,  
 Kanagawa 230-0045, Japan (E-mail: genome-res@gsr.riken.jp,  
 URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222,  
 Fax: 81-45-503-9216)

## COMMENT

cDNA library was prepared and sequenced in Mouse Genome  
 Encyclopedia Project of Genome Exploration Research Group in Riken  
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
 Division of Experimental Animal Research in Riken contributed to  
 prepare mouse tissues.  
 Please visit our web site for further details.  
 URL: http://genome.gsc.riken.jp/  
 URL: http://fantom.gsc.riken.jp/  
 Location/Qualifiers

## FEATURES

## source

1..812  
 /organism="Mus musculus"  
 /mol\_type="RNA"  
 /strain="CS7BL/6J"  
 /db\_xref="FANTOM:DB:4833404E21"  
 /db\_xref="taxon:10090"  
 /clone="4833404E21"  
 /issue\_type="head"  
 /clone\_lib="RIKEN full-length enriched mouse cDNA library"  
 /dev\_stage="0 day neonate"  
 73..663  
 /note="unnamed protein product; amelogenin (MGD|MG1:88005,  
 GB|D31768, evidence: BLASTN, 99%, match=799)  
 putative"  
 /codon\_start=1  
 /protein\_id="BAC26415.1"  
 /db\_xref="GI:26325322"  
 /translation="MGTWLPRACLLGAFAFPLPPHSGSPGYINLSYEVLTPLKMYQS  
 MIRQYPSYGYEPMGWLHQLIPVLSQHPSPHTLQPHNLPPVPAQAPVAPQPPM  
 PVPGHSMPTPDHQPNIIPSAQQPFPQPAIPQSHQPMQFQSPDLHMQPLAPQ  
 PPLPPLFPMQPLSPILPELPLEAMPATDKTKEEVD"  
 793..798  
 /note="putative"  
 812  
 /note="putative"

## CDS

polyA\_signal  
 polyA\_site

## ORIGIN

Query Match 100.0%; Score 22; DB 3; Length 812;  
 Best Local Similarity 100.0%; Pred. No. 32;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GATTGTGTTGCTGCTGCTGCTG 22  
 84 GATTGTGTTGCTGCTGCTGCTG 105

## Db

RESULT 8  
 CB588212 843 bp mRNA linear EST 03-APR-2003  
 LOCUS CB588212  
 DEFINITION AGENCOURT 12771359 NIH MGC 136 Mus musculus cDNA clone  
 accession CB588212  
 VERSION CB588212.1 GI:29506068  
 KEYWORDS EST.



SOURCE  
ORGANISM Mus musculus (house mouse)

REFERENCE  
AUTHORS Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
TITLE NIH-MGC http://mgc.nci.nih.gov/.  
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)  
COMMENT Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-remail.nih.gov  
Tissue Procurement: Dr. David Rowe  
CDNA Library Preparation: Invitrogen Corp  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: NDAM318 row: 9 column: 08  
High quality sequence stop: 636.  
Location/Qualifiers  
1. 843  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:30293239"  
/issue\_type="embryonic limb, maxilla and mandible"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_136"  
/note="Vector: PCMV-SPORT6.1; Site 1: EcoRV; Site 2: NotI;  
Normalized, full-length enriched library from pool of  
mouse embryonic limb, maxilla and mandible, embryonic day  
17.5, 18.5 and newborn (mandible (5' 4 and 1 limb and jaw  
equivalents from respective days). Cloned directionally,  
oligo-dT primed (5'-GACTAGTCTAGATCGAGCGGCCGCC(T)15-3'.  
Size selected for the >1kb fragments, average insert size  
1.2 kb. Normalization to Cot 7.5. Tissue contributed by  
David Rowe; library constructed by Resgen, Invitrogen  
Corp. Note: this is a NIH\_MGC Library."

ORIGIN

Query Match 100.0%; Score 22; DB 6; Length 843;  
Best Local Similarity 100.0%; Pred. No. 32;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GATTTGTTGGCTGCTCTCTG 22  
|||||  
Db 51 GATTTGTTGGCTGCTCTCTG 72  
|||||

RESULT 9  
CB588525 859 bp mRNA linear EST 03-APR-2003  
LOCUS AGENCOURT 12567944 NIH MGC 136 Mus musculus CDNA clone  
DEFINITION IMAGE:30289751 5', mRNA sequence.  
ACCESSION CB588525  
VERSION CB588525.1 GI:29506381  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Mammalia; Euteleostomi; Chordata; Craniata; Vertebrata; Euteleostomi;  
TITLE NIH-MGC http://mgc.nci.nih.gov/.  
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)  
COMMENT Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-remail.nih.gov  
Tissue Procurement: Dr. David Rowe  
CDNA Library Preparation: Invitrogen Corp  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov  
Plate: NDAM329 row: e column: 24  
High quality sequence stop: 763.  
Location/Qualifiers  
1. 859  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:30289751"  
/issue\_type="embryonic limb, maxilla and mandible"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_136"  
/note="Vector: PCMV-SPORT6.1; Site 1: EcoRV; Site 2: NotI;  
Normalized, full-length enriched library from pool of  
mouse embryonic limb, maxilla and mandible, embryonic day  
17.5, 18.5 and newborn (mandible (5' 4 and 1 limb and jaw  
equivalents from respective days). Cloned directionally,  
oligo-dT primed (5'-GACTAGTCTAGATCGAGCGGCCGCC(T)15-3'.  
Size selected for the >1kb fragments, average insert size  
1.2 kb. Normalization to Cot 7.5. Tissue contributed by  
David Rowe; library constructed by Resgen, Invitrogen  
Corp. Note: this is a NIH\_MGC Library."

ORIGIN

Query Match 100.0%; Score 22; DB 6; Length 859;  
Best Local Similarity 100.0%; Pred. No. 32;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GATTTGTTGGCTGCTCTCTG 22  
|||||  
Db 70 GATTTGTTGGCTGCTCTCTG 91  
|||||

RESULT 10  
CB589177 865 bp mRNA linear EST 03-APR-2003  
LOCUS AGENCOURT 12770446 NIH MGC 136 Mus musculus CDNA clone  
DEFINITION IMAGE:30290534 5', mRNA sequence.  
ACCESSION CB589177  
VERSION CB589177.1 GI:29507033  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
TITLE NIH-MGC http://mgc.nci.nih.gov/.  
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)  
COMMENT Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-remail.nih.gov  
Tissue Procurement: Dr. David Rowe  
CDNA Library Preparation: Invitrogen Corp  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: NDAM331 row: f column: 15  
High quality sequence stop: 630.  
Location/Qualifiers  
1. 865  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:30290534"  
/issue\_type="embryonic limb, maxilla and mandible"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_136"  
/note="Vector: PCMV-SPORT6.1; Site 1: EcoRV; Site 2: NotI;  
Normalized, full-length enriched library from pool of  
mouse embryonic limb, maxilla and mandible, embryonic day  
17.5, 18.5 and newborn (mandible (5' 4 and 1 limb and jaw

equivalents from respective days). Cloned directionally, oligo-dT primed (5'-GACTAGTTCTAGATCGGACGCGCCCTT)15-3'. Size selected for the >1kb fragments, average insert size 1.2 kb. Normalization to Cot 7.5. Tissue contributed by David Rowe; library constructed by Resgen, Invitrogen Corp. Note: this is a NIH\_MGC library."

## ORIGIN

Query Match 100.0%; Score 22; DB 6; Length 865;  
Best Local Similarity 100.0%; Pred. No. 32;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 GATTTGTTGCTGCTGCTGCTG 22  
|||||  
Db 69 GATTTGTTGCTGCTGCTGCTG 90

RESULT 11  
LOCUS CB587051 886 bp mRNA linear EST 03-APR-2003  
DEFINITION AGENCOURT 12807074 NIH\_MGC\_136 Mus musculus cDNA clone  
IMAGE:30295123 5', mRNA sequence.  
ACCESSION CB587051  
VERSION CB587051.1 GI:29504907  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 886)  
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: Dr. David Rowe  
CDNA Library Preparation: Invitrogen Corp  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>  
Plate: NDAM343 row: e column: 20  
High quality sequence stop: 619.  
Location/Qualifiers

## FEATURES

1..886  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:30295123"  
/tissue\_type="embryonic limb, maxilla and mandible"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_1ib="NIH\_MGC\_136"  
/note="Vector: PCMV-SPORT6.1; Site 1: EcoRV; Site 2: NotI; Normalized, full-length enriched library from pool of mouse embryonic limb, maxilla and mandible, embryonic day 17.5, 18.5 and newborn (mandible (5, 4 and 1 limb and jaw equivalents from respective days). Cloned directionally, oligo-dT primed (5'-GACTAGTTCTAGATCGGACGCGCCCTT)15-3'. Size selected for the >1kb fragments, average insert size 1.2 kb. Normalization to Cot 7.5. Tissue contributed by David Rowe; library constructed by Resgen, Invitrogen Corp. Note: this is a NIH\_MGC library."

## ORIGIN

Query Match 100.0%; Score 22; DB 6; Length 886;  
Best Local Similarity 100.0%; Pred. No. 32;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 GATTTGTTGCTGCTGCTGCTG 22  
|||||  
Db 95 GATTTGTTGCTGCTGCTGCTG 116

RESULT 12  
LOCUS CB589251 889 bp mRNA linear EST 03-APR-2003  
DEFINITION AGENCOURT 12760700 NIH\_MGC\_136 Mus musculus cDNA clone  
IMAGE:30287069 5', mRNA sequence.

ACCESSION CB589251  
VERSION CB589251.1 GI:29507107  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 889)  
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: Dr. David Rowe  
CDNA Library Preparation: Invitrogen Corp  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>  
Plate: NDAM322 row: f column: 06  
High quality sequence start: 23  
High quality sequence stop: 581.  
Location/Qualifiers

## FEATURES

1..889  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:30287069"  
/tissue\_type="embryonic limb, maxilla and mandible"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_1ib="NIH\_MGC\_136"  
/note="Vector: PCMV-SPORT6.1; Site 1: EcoRV; Site 2: NotI; Normalized, full-length enriched library from pool of mouse embryonic limb, maxilla and mandible, embryonic day 17.5, 18.5 and newborn (mandible (5, 4 and 1 limb and jaw equivalents from respective days). Cloned directionally, oligo-dT primed (5'-GACTAGTTCTAGATCGGACGCGCCCTT)15-3'. Size selected for the >1kb fragments, average insert size 1.2 kb. Normalization to Cot 7.5. Tissue contributed by David Rowe; library constructed by Resgen, Invitrogen Corp. Note: this is a NIH\_MGC library."

## ORIGIN

Query Match 100.0%; Score 22; DB 6; Length 889;  
Best Local Similarity 100.0%; Pred. No. 32;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 GATTTGTTGCTGCTGCTGCTG 22  
|||||  
Db 89 GATTTGTTGCTGCTGCTGCTG 110

RESULT 13  
LOCUS CB574837 891 bp mRNA linear EST 02-APR-2003  
DEFINITION AGENCOURT 12973279 NIH\_MGC\_136 Mus musculus cDNA clone  
IMAGE:30294719 5', mRNA sequence.  
ACCESSION CB574837  
VERSION CB574837.1 GI:29494367  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 891)  
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE  
JOURNAL  
COMMENT  
National Instructors of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-remail.nih.gov  
Tissue Procurement: Dr. David Rowe  
CDNA Library Preparation: Invitrogen Corp  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:  
http://image.llnl.gov  
Plate: NDAM342 row: d column: 24  
High quality sequence start: 19  
High quality sequence stop: 658.  
Location/Qualifiers

## FEATURES

source

1. .891  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:30294719"  
/tissue\_type="embryonic limb, maxilla and mandible"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_1lb="NIH\_MGC\_136"  
/note="Vector: PCMV-SPORE.1; Site 1: EcoRV; Site 2: NotI; Normalized, full-length enriched library from pool of mouse embryonic limb, maxilla and mandible, embryonic day 17.5, 18.5 and newborn (mandible (5, 4 and 1 limb and jaw equivalents from respective days). Cloned directionally, oligo-dT primed (5'-GACTAGTCTAGATCGGCGGCCGCC(T)15-3', size selected for the >1kb fragments, average insert size 1.2 kb. Normalization to Cot 7.5. Tissue contributed by David Rowe; library constructed by Resgen, Invitrogen Corp. Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 100.0%; Score 22; DB 6; Length 891;  
Best Local Similarity 100.0%; Pred. No. 32;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATTTGTTGGCTGCTCCTG 22  
|||||  
DB 90 GATTTGTTGGCTGCTCCTG 111

RESULT 14  
CB590451 913 bp mRNA linear EST 03-APR-2003  
LOCUS  
DEFINITION  
AGENCOURT 12770290 NIH\_MGC\_136 Mus musculus CDNA clone  
IMAGE:30286944 5', mRNA sequence.  
CB590451  
ACCESSION  
CB590451.1 GI:29508307  
VERSION  
KEYWORDS  
EST.  
SOURCE  
Mus musculus (house mouse)  
ORGANISM  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 913)  
NIH-MGC http://mgc.nci.nih.gov/  
National Instructors of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-remail.nih.gov  
Tissue Procurement: Dr. David Rowe  
CDNA Library Preparation: Invitrogen Corp  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:  
http://image.llnl.gov  
Plate: NDAM322 row: a column: 01  
High quality sequence start: 12  
High quality sequence stop: 640.  
Location/Qualifiers

## FEATURES

location/Qualifiers

source

1. .913  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:30286944"  
/tissue\_type="embryonic limb, maxilla and mandible"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_1lb="NIH\_MGC\_136"  
/note="Vector: PCMV-SPORE.1; Site 1: EcoRV; Site 2: NotI; Normalized, full-length enriched library from pool of mouse embryonic limb, maxilla and mandible, embryonic day 17.5, 18.5 and newborn (mandible (5, 4 and 1 limb and jaw equivalents from respective days). Cloned directionally, oligo-dT primed (5'-GACTAGTCTAGATCGGCGGCCGCC(T)15-3', size selected for the >1kb fragments, average insert size 1.2 kb. Normalization to Cot 7.5. Tissue contributed by David Rowe; library constructed by Resgen, Invitrogen Corp. Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 100.0%; Score 22; DB 6; Length 913;  
Best Local Similarity 100.0%; Pred. No. 32;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATTTGTTGGCTGCTCCTG 22  
|||||  
DB 82 GATTTGTTGGCTGCTCCTG 103

RESULT 15  
CB587332 919 bp mRNA linear EST 03-APR-2003  
LOCUS  
DEFINITION  
AGENCOURT 12971606 NIH\_MGC\_136 Mus musculus CDNA clone  
IMAGE:30291971 5', mRNA sequence.  
CB587332  
ACCESSION  
CB587332.1 GI:29505188  
VERSION  
KEYWORDS  
EST.  
SOURCE  
Mus musculus (house mouse)  
ORGANISM  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 919)  
NIH-MGC http://mgc.nci.nih.gov/  
National Instructors of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-remail.nih.gov  
Tissue Procurement: Dr. David Rowe  
CDNA Library Preparation: Invitrogen Corp  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:  
http://image.llnl.gov  
Plate: NDAM335 row: b column: 12  
High quality sequence start: 216  
High quality sequence stop: 325.  
Location/Qualifiers

## FEATURES

1. .919  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:30291971"  
/tissue\_type="embryonic limb, maxilla and mandible"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_1lb="NIH\_MGC\_136"  
/note="Vector: PCMV-SPORE.1; Site 1: EcoRV; Site 2: NotI; Normalized, full-length enriched library from pool of mouse embryonic limb, maxilla and mandible, embryonic day 17.5, 18.5 and newborn (mandible (5, 4 and 1 limb and jaw equivalents from respective days). Cloned directionally, oligo-dT primed (5'-GACTAGTCTAGATCGGCGGCCGCC(T)15-3', size selected for the >1kb fragments, average insert size

Size selected for the >1kb fragments, average insert size

1.2 kb. Normalization to Cot 7.5. Tissue contributed by  
David Rowe; library constructed by Resgen, Invitrogen  
Corp. Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 100.0%; Score 22; DB 6; Length 919;  
Best Local Similarity 100.0%; Pred. No. 32;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GATTTGTTGGCTGCTCTCTG 22  
|||  
Db 69 GATTTGTTGGCTGCTCTCTG 90

Search completed: February 8, 2005, 22:17:57  
Job time : 267.041 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 8, 2005, 11:28:00 ; Search time 122.665 Seconds  
(without alignments)  
8481.433 Million cell updates/sec

Title: US-10-754-437-5  
Perfect score: 22  
Sequence: 1 gactctgtactgtcttaacc 22

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4526729 seqs, 23644849745 residues  
Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmb1:\*  
1: gb\_ba:\*  
2: gb\_btg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sbs:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	18.8	85.5	6451	4 AB091789	AB091789 Bos tauri
2	18.8	85.5	208214	10 AL645908	AL645908 Mouse DNA
3	18.4	83.6	184836	2 AC140360	AC140360 Mus muscu
4	18.4	81.8	162927	10 AC115302	AC115302 Mus muscu
5	18.8	81.8	209617	10 AC139300	AC139300 Mus muscu
6	17.8	80.9	108799	9 AC090989	AC090989 Homo sapi
7	17.8	80.9	168637	5 AL844187	AL844187 Zebrafish
8	17.8	80.9	176148	9 AC010798	AC010798 Homo sapi
9	17.8	80.9	177151	2 AC132312	AC132312 Mus muscu
10	17.8	80.9	178139	2 AC118710	AC118710 Mus muscu
11	17.8	80.9	178825	2 AP001203	AP001203 Homo sapi
12	17.8	80.9	188972	10 AC101658	AC101658 Homo sapi
13	17.8	80.9	208244	2 BX470182	BX470182 Danio rer
14	17.8	80.9	213071	2 AC110680	AC110680 Rattus no
15	17.8	80.9	223538	2 AC087150	AC087150 Mus muscu
16	17.8	80.9	228660	2 AC132354	AC132354 Mus muscu
17	17.8	80.9	244868	2 AC095331	AC095331 Rattus no
18	17.8	80.9	245528	2 AC133363	AC133363 Rattus no
19	17.8	80.9	251219	2 CR388422	CR388422 Danio rer

#### ALIGNMENTS

RESULT 1	AB091789/c	6451 bp	DNA	linear	MAM 02-MAY-2003
LOCUS	Bos taurus	AMELX gene for amelogenin, partial cds.			
DEFINITION	AB091789				
ACCESSION	AB091789.1	GI:29126030			
VERSION					
KEYWORDS					
SOURCE	Bos taurus (cow)				
ORGANISM	Bos taurus				
REFERENCE	Iwase,M., Satte,Y., Hirai,Y., Hirai,H., Imai,H. and Takahata,N.				
AUTHORS	From the Cover: The amelogenin loci span an ancient pseudautosomal boundary in diverse mammalian species				
TITLE	Proc. Natl. Acad. Sci. U.S.A. 100 (9), 5258-5263 (2003)				
JOURNAL	22608569				
MEDLINE	12672962				
PUBMED	2 (bases 1 to 6451)				
REFERENCE	Iwase,M., Satte,Y. and Takahata,N.				
AUTHORS	Submitted (19-SEP-2002) Mineyo Iwase, Graduate University for Advanced Studies(Sokendai), Department of Biosystems Science; Shoman Kokusaiinura, Hayama, Kanagawa 240-0193, Japan (E-mail:iwase@min@koryu.w01.soken.ac.jp, Tel:81-468-58-1571, Fax:81-468-58-1544)				
TITLE	Location/Qualifiers				
JOURNAL	1..6451				
FEATURES	/organism="Bos taurus"				
source	/mol_type="genomic DNA"				
	/db_xref="taxon:9913"				
	/sex="male"				
gene	238..6451				
	/gene="AMELX"				
5'UTR	join(238..340,1650..1661)				
	/gene="AMELX"				
	join(1662..1715,4173..4220,5560..5601,5665..5709,5987..6451)				
CDS	/gene="AMELX"				
	/codon_start=1				
	/product="amelogenin"				

20	17.8	80.9	270721	2 AC105775	AC105775 Rattus no
21	17.4	79.1	71701	2 AF254984	AF254984 Homo sapi
22	17.4	79.1	75272	2 AC136287	AC136287 Medicago
23	17.4	79.1	86554	2 AC090280	AC090280 Homo sapi
24	17.4	79.1	105191	9 AL161655	AL161655 Human DNA
25	17.4	79.1	122924	2 AF207954	AF207954 Homo sapi
26	17.4	79.1	124630	9 AC090731	AC090731 Homo sapi
27	17.4	79.1	146734	10 AL590415	AL590415 Mouse DNA
28	17.4	79.1	167232	5 AL929379	AL929379 Zebrafish
29	17.4	79.1	172875	9 AC084301	AC084301 Homo sapi
30	17.4	79.1	195119	5 BX649594	BX649594 Zebrafish
31	17.4	79.1	214811	10 AL662895	AL662895 Mouse DNA
32	17.4	79.1	215615	2 AC098139	AC098139 Rattus no
33	17.4	79.1	326786	2 AC022365	AC022365 Homo sapi
34	17.2	78.2	334	8 AY142645	AY142645 Arabidops
35	17.2	78.2	363	6 BD248010	BD248010 5' ESTs f
36	17.2	78.2	363	6 CO771973	CO771973 Sequence
37	17.2	78.2	397	11 HS2822E1	Z66933 H.sapiens D
38	17.2	78.2	479	6 AR415985	AR415985 Sequence
39	17.2	78.2	479	6 AX972819	AX972819 Sequence
40	17.2	78.2	479	6 BD111538	BD111538 EST and e
41	17.2	78.2	488	8 AY034985	AY034985 Arabidops
42	17.2	78.2	640	8 AY086566	AY086566 Arabidops
43	17.2	78.2	1165	9 BC000735	BC000735 Homo sapi
44	17.2	78.2	1645	9 HSY13613	YS13613 Homo sapien
45	17.2	78.2	1935	9 D83730	D83730 Homo sapien

/protein\_id="BAC66108.1"  
 /db\_xref="gi:29126031"  
 /translation="MGTRITLACILGAAPSPLEPHGPGYINFEYSNYPGISTD  
 ETALVLPKMTQSMIRHYPISYIEPMGWLHHQIIPVSOQTPQNHAIQPHHILPM  
 VPAQGVVPOQPMKVPQSHMPTQHHPMLPLPAQOPPOPSICPQHPQPLQHPQ  
 LQPMQMPLOPLQLOPQFVHPIQPLPQPPPLPPIFMQPLPPLPDLPLAMPPR"

ORIGIN  
 Query Match 85.5%; Score 18.8; DB 4; Length 6451;  
 Best Local Similarity 90.9%; Pred. No. 1.1e+02;  
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 Qy 1 GATCTGTACTGTCTTAAACC 22  
 |||||  
 Db 1821 GATCATGTACTGTCTTAAACC 1800

RESULT 2  
 AL645908 208214 bp. DNA linear ROD 05-APR-2002  
 LOCUS Mouse DNA sequence from clone RP23-24A24 on chromosome 11, complete  
 DEFINITION sequence.  
 ACCESSION AL645908 GI:20068640  
 VERSION AL645908.9  
 KEYWORDS HTG.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus

REFERENCE  
 AUTHORS Pearce, A.  
 TITLE Direct Submission  
 JOURNAL Submitted (04-APR-2002) Wellcome Trust Sanger Institute, Hinxton,  
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
 humquerry@sanger.ac.uk; Clone requests: clonerequest@sanger.ac.uk

## COMMENT

On Apr 7, 2002 this sequence version replaced gi:17976640.  
 During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variation annotation may not be found in the sequence submission  
 corresponding to the overlapping clone, as we submit sequences with  
 only a small overlap as described above.  
 This sequence was finished as follows unless otherwise noted: all  
 regions were either double-stranded or sequenced with an alternate  
 chemistry or covered by high quality data (i.e., phred quality >=  
 30); an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by at least  
 one plasmid subclone or more than one M13 subclone; and the  
 assembly was confirmed by restriction digest. The following  
 abbreviations are used to associate primary accession numbers given  
 in the feature table with their source databases: Em; EMBL; Sw;  
 SWISSPROT; Tr; TREMBL; Wp; WORMPEP; Information on the WORMPEP  
 database can be found at  
 http://www.sanger.ac.uk/Projects/C\_elegans/wormpep  
 from the RPI-23 Mouse PAC library  
 constructed by the group of Pieter de Jong.  
 For further details see http://www.chori.org/bacpac/home.htm  
 VECTOR: pBACe3.6.

## FEATURES

source  
 1..208214  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10090"  
 /chromosome="11"  
 /clone="RP23-24A24"  
 /clone\_1fb="RPCT-23"

## ORIGIN

Query Match 85.5%; Score 18.8; DB 10; Length 208214;  
 Best Local Similarity 90.9%; Pred. No. 66;  
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GATCTGTACTGTCTTAAACC 22  
 |||||  
 Db 80649 GATCTGTACTGTCTTAAACC 80670

RESULT 3  
 AC140360/c 184836 bp DNA linear HTG 12-JUN-2004  
 LOCUS Mus musculus chromosome 12 clone RP24-24IG4, WORKING DRAFT  
 DEFINITION SEQUENCE, 5 unordered pieces.  
 ACCESSION AC140360  
 VERSION AC140360.2 GI:48675488  
 KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ACTIVEFIN.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus

REFERENCE  
 AUTHORS Wilson, R.K.  
 TITLE The sequence of Mus musculus clone  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 184836)  
 AUTHORS McPherson, J.D. and Waterston, R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (23-FEB-2003) Genome Sequencing Center, 4444 Forest Park  
 Parkway, St. Louis, MO 63108, USA  
 REFERENCE 3 (bases 1 to 184836)  
 AUTHORS Wilson, R.K.  
 TITLE Direct Submission  
 JOURNAL Submitted (12-JUN-2004) Genome Sequencing Center, 4444 Forest Park  
 Parkway, St. Louis, MO 63108, USA  
 On Jun 12, 2004 this sequence version replaced gi:28475618.

## COMMENT

----- Genome Center -----  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: http://genome.wustl.edu  
 Contact: submissions@wustl.wustl.edu  
 Project Information -----  
 Center project name: W\_BB0241G04  
 ----- Summary Statistics -----  
 Sequencing vector: M13, 0%  
 Sequencing vector: plasmid, 100%  
 Chemistry: Dye-primer ET, 0% of reads  
 Chemistry: Dye-terminator Big Dye, 100% of reads  
 Assembly program: Phrap; version 0.990319  
 Consensus quality: 182119 bases at least Q40  
 Consensus quality: 182305 bases at least Q30  
 Consensus quality: 182480 bases at least Q20

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 5 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 4967: contig of 4967 bp in length  
 \* 4968 5067: gap of unknown length  
 \* 5068 12306: contig of 7239 bp in length  
 \* 12307 12406: gap of unknown length  
 \* 12407 35592: contig of 23186 bp in length  
 \* 35593 35692: gap of unknown length  
 \* 35693 88170: contig of 52478 bp in length  
 \* 88171 88271: gap of unknown length  
 \* 88271 184836: contig of 96566 bp in length.

## FEATURES

source  
 1..184836  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10090"

```

/misc_feature      /chromosome="12"
                   /clone="RP24-241G4"
                   1..4967
/misc_feature      /note="assembly_name:Contig51"
                   5068..12306
/misc_feature      /note="assembly_name:Contig52"
                   12407..35592
/misc_feature      /note="assembly_name:Contig53"
                   35693..88170
/misc_feature      /note="assembly_name:Contig54"
                   88271..184836
/misc_feature      /note="assembly_name:Contig55"

```

## ORIGIN

```

Query Match      83.6%; Score 18.4; DB 2; Length 184836;
Best Local Similarity 95.0%; Pred. No. 1.1e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 3 TCTTGACTGCTTAACCC 22
      |||||
Db 408 TCTTGACTGCTTAACCC 389

```

```

RESULT 4
AC115302      162927 bp  DNA  linear  ROD 27-NOV-2003
LOCUS      Mus musculus BAC clone RP24-107D24 from chromosome 18, complete
AC115302      AC115302      GI:30142728
VERSION      HTG.
KEYWORDS     Mus musculus (house mouse)
SOURCE       Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
              1 (bases 1 to 162927)
              Boatright, E., Walligorski, J., Haakenson, W. and Doeber, A.
              The sequence of Mus musculus BAC clone RP24-107D24
              Unpublished (2001)
              2 (bases 1 to 162927)
              Wilson, R.
              Sequencing of Mus musculus
              Unpublished (2001)
              3 (bases 1 to 162927)
              McPherson, J.D. and Waterston, R.H.
              Direct Submision
              Submitted (16-MAR-2002) Genome Sequencing Center, 4444 Forest Park
              Parkway, St. Louis, MO 63108, USA
              4 (bases 1 to 162927)
              McPherson, J.D. and Waterston, R.H.
              Direct Submision
              Submitted (25-AUG-2002) Genome Sequencing Center, 4444 Forest Park
              Parkway, St. Louis, MO 63108, USA
              5 (bases 1 to 162927)
              McPherson, J.D. and Waterston, R.H.
              Direct Submision
              Submitted (27-APR-2003) Genome Sequencing Center, 4444 Forest Park
              Parkway, St. Louis, MO 63108, USA
              6 (bases 1 to 162927)
              Wilson, R.
              Direct Submision
              Submitted (27-NOV-2003) Department of Genetics, Washington
              University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
              On Apr 27, 2003 this sequence version replaced gi:22475848.

```

## COMMENT

```

-----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@wustl.wustl.edu
-----
Summary Statistics
-----
Center project name: M_BB0107D24
-----

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

## MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu>

## SOURCE INFORMATION:

The RPCI-24 BAC Library has been constructed by Pieter de Jong and coworkers (<http://www.chori.org>) from male C57BL/6J mouse spleen and/or brain genomic DNA. The clone and detailed information can be obtained from Pieter de Jong and coworkers at <http://www.chori.org>

## NEIGHBORING SEQUENCE INFORMATION:

This sequence is the entire insert of the clone. This clone is overlapped by AC131796.

## FEATURES

## source

```

1..162927
Location/Qualifiers

```

```

/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="18"
/map="18"

```

```

/clone="RP24-107D24"
/clone_1lb="RPCI-24"

```

```

1..1209

```

```

/rpc_family="L1"

```

```

/misc_feature
/note="Sequence derived from PCR product of project DNA."

```

```

1209..1556

```

```

/rpc_family="L1"

```

```

1617..1938

```

```

/rpc_family="MaLR"

```

```

1941..2463

```

```

/rpc_family="MaLR"

```

```

2506..3065

```

```

/rpc_family="MaLR"

```

```

3070..3209

```

```

/rpc_family="Alu"

```

```

3504..3754

```

```

/rpc_family="MaLR"

```

```

3755..4362

```

```

/rpc_family="ERV1"

```

```

4363..5160

```

```

/rpc_family="ERV1"

```

```

5164..5315

```

```

/rpc_family="ERV1"

```

```

8477..9064

```

```

/rpc_family="ERV1"

```

```

9067..9644

```

```

/rpc_family="MaLR"

```

```

9645..9970

```

```

/rpc_family="MaLR"

```

```

10028..10128

```

```

/rpc_family="B1"

```

```

10128..10512

```

```

/rpc_family="B1"

```

```

10642..10721

```

```

/rpc_family="B1"

```

```

10870..11144

```

```

/rpc_family="B1"

```

```

repeat_region 11752..14937 /rpt_family="L1"
repeat_region 14938..15072 /rpt_family="ERV1"
repeat_region 18731..19077 /rpt_family="MALR"
repeat_region 19931..20167 /rpt_family="L1"
repeat_region 20151..20262 /rpt_family="L1"
repeat_region 20309..20620 /rpt_family="L1"
repeat_region 20727..22087 /rpt_family="L1"
repeat_region 22105..22200 /rpt_family="L1"
repeat_region 22348..22642 /rpt_family="L1"
repeat_region 22846..22991 /rpt_family="RMR19A"
repeat_region 23027..23427 /rpt_family="L1"
repeat_region 23410..23734 /rpt_family="L1"
repeat_region 26778..27659 /rpt_family="L1"
repeat_region 28064..28180 /rpt_family="L1"
repeat_region 28179..28783 /rpt_family="RMR13B"
repeat_region 28558..28794 /rpt_family="RMR13B"
repeat_region 29543..29690 /rpt_family="ERVK"
repeat_region 29723..30108 /rpt_family="B4"
repeat_region 30799..30977 /rpt_family="MALR"
repeat_region 32579..32823 /rpt_family="ERV1"
repeat_region 33660..33807 /rpt_family="L1"
repeat_region 34460..34606 /rpt_family="Alu"
repeat_region 34639..34846 /rpt_family="Alu"
repeat_region 35192..35438 /rpt_family="B2"
repeat_region 35703..35938 /rpt_family="MALR"
repeat_region 35993..36399 /rpt_family="MALR"
repeat_region 37273..37677 /rpt_family="MALR"
repeat_region 38289..38533 /rpt_family="MALR"
repeat_region 38542..39278 /rpt_family="MALR"
repeat_region 39297..40291 /rpt_family="L1"
repeat_region 40292..40668 /rpt_family="L1"
repeat_region 40669..40815 /rpt_family="MALR"
repeat_region 40816..40896 /rpt_family="Alu"
repeat_region 40897..41065 /rpt_family="MALR"
repeat_region 41715..41994 /rpt_family="L1"
repeat_region 42350..42542 /rpt_family="L1"
repeat_region 43595..43724 /rpt_family="B2"

```

```

Query Match 81.8%; Score 18; DB 10; Length 162927;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 134116 ATCTGTACTGCTTAA 134133

2 ATCTGTACTGCTTAA 19
|||||
Db

RESULT 5
AC139300/c 209617 bp DNA linear ROD 25-NOV-2003
LOCUS Mus musculus BAC clone RP23-258L8 from chromosome 18, complete
DEFINITION
AC139300
AC139300 GI:37361067
VERSION
AC139300.4
KEYWORDS
HTG.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 209617)
Levy, A.
The sequence of Mus musculus BAC clone RP23-258L8
Unpublished (2001)
2 (bases 1 to 209617)
Wilson, R.
Sequencing of Mus musculus
Unpublished (2001)
3 (bases 1 to 209617)
McPherson, J.D. and Waterston, R.H.
Direct Submission
Submitted (29-JAN-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
4 (bases 1 to 209617)
Wilson, R.K.
Direct Submission
Submitted (20-MAY-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
5 (bases 1 to 209617)
Wilson, R.K.
Direct Submission
Submitted (02-OCT-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
6 (bases 1 to 209617)
Wilson, R.
Direct Submission
Submitted (25-NOV-2003) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Oct 2, 2003 this sequence version replaced gi:30911162.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@wustl.wustl.edu
----- Summary Statistics
Center project name: M_BA0258L08

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate







```

repeat_region      /rpt_family="MADE1"
                    17547..17567
repeat_region      /rpt_family="AT_r1ch"
                    complement(18293..18508)
repeat_region      /rpt_family="MLT1J"
                    18941..19018
repeat_region      /rpt_family="MIR"
                    19077..19098
repeat_region      /rpt_family="AT_r1ch"
                    19453..19756
repeat_region      /rpt_family="AluJo"
                    20661..20872
repeat_region      /rpt_family="L2"
                    complement(23626..23836)
repeat_region      /rpt_family="MIR"
                    25418..25458
repeat_region      /rpt_family="AT_r1ch"
                    27001..27112
repeat_region      /rpt_family="LIME"
                    complement(27235..27408)
repeat_region      /rpt_family="MERSB"
                    27508..27533
repeat_region      /rpt_family="AT_r1ch"
                    29143..29441
repeat_region      /rpt_family="AluSc"
                    complement(30072..30333)
repeat_region      /rpt_family="LIMC/D"
                    complement(30971..31057)
repeat_region      /rpt_family="MLT1K"
                    31882..31946
repeat_region      /rpt_family="(CAA)n"
                    complement(32318..32545)
repeat_region      /rpt_family="MER46A"
                    33184..33214
repeat_region      /rpt_family="AT_r1ch"
                    33562..33613
repeat_region      /rpt_family="L2"
                    33682..33816
repeat_region      /rpt_family="(TA)n"
                    complement(33817..33938)
repeat_region      /rpt_family="FLAM_C"
                    34204..34421
repeat_region      /rpt_family="MIR"
                    34576..34607
repeat_region      /rpt_family="AT_r1ch"
                    complement(34716..34944)
repeat_region      /rpt_family="MIR"
                    35815..35875
repeat_region      /rpt_family="(TA)n"
                    complement(35975..35979)
unrepeated          /note="30 qual SNGI region"
repeat_region      36347..36377
repeat_region      /rpt_family="AT_r1ch"
                    36547..36611
repeat_region      /rpt_family="AT_r1ch"
                    37724..37744
repeat_region      /rpt_family="AT_r1ch"
                    38576..38603
repeat_region      /rpt_family="AT_r1ch"
                    40354..40383
repeat_region      /rpt_family="AT_r1ch"
                    41112..41139
repeat_region      /rpt_family="AT_r1ch"
                    complement(43246..46180)
repeat_region      /rpt_family="L2"

```

```

Query Match      80.9%; Score 17.8; DB 9; Length 108799;
Best Local Similarity 90.5%; Pred. No. 2.3e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY      2 ATCTGTACTGTCTTAACCC 22
DB      84043 ATCTGTACTGTCTTAACCC 84023

```

```

RESULT 7
AL844187          168637 bp      DNA      linear      VRT 13-DEC-2003
LOCUS            Zebrafish DNA sequence from clone DKEY-221D18 in linkage group 16,
DEFINITION       complete sequence.
ACCESSION        AL844187.13 GI:39919157
VERSION          HTG.
KEYWORDS         Danio rerio (zebrafish)
SOURCE           Danio rerio
ORGANISM         Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                  Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
                  Cypriniformes; Cyprinidae; Danio.
REFERENCE        Submitted (13-DEC-2003) Wellcome Trust Sanger Institute, Hinxton,
AUTHORS          Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
TITLE            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
JOURNAL          On Dec 15, 2003 this sequence version replaced gi:29119912.
COMMENT          ----- Genome Center
                  Center: Wellcome Trust Sanger Institute
                  Center code: SC
                  Web site: http://www.sanger.ac.uk
                  Contact: zfish-help@sanger.ac.uk
                  -----

```

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; SW: SWISSPROT; Tr: TRMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) DKEY-221D18 is from a Zebrafish BAC library

VECTOR: pindigBAC-5

Clone-derived Zebrafish pUC subclones occasionally display inconsistency over the length of mononucleotide A/T runs and conserved TA repeats. Where this is found the longest good quality representation will be submitted.

Repeat names beginning 'Dr' were identified by the Recon repeat discovery system (Zhirong Bao and Sean Eddy, submitted), and those beginning 'dir' were identified by Rick Waterman (Stephen Johnson lab, WashU). For further information see [http://www.sanger.ac.uk/Projects/D\\_rerio/fishmark.shtml](http://www.sanger.ac.uk/Projects/D_rerio/fishmark.shtml).

## FEATURES

```

source
1..168637
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="DKEY-221D18"
/clone_1lb="Dantiokey"

```

## ORIGIN

```

Query Match      80.9%; Score 17.8; DB 5; Length 168637;
Best Local Similarity 90.5%; Pred. No. 2.2e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY      2 ATCTGTACTGTCTTAACCC 22

```





Query Match 80.9%; Score 17.8; DB 2; Length 177151;  
 Best Local Similarity 90.5%; Pred. No. 2.2e+02;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GATCTTGACTGCTTAACC 21  
 |||||  
 DB 111099 GGTCTGTACTGTCTTAAGCC 111119

RESULT 10  
 AC118710 178139 bp DNA linear HTG 29-APR-2004  
 LOCUS Mus musculus chromosome 15 clone RP24-121C9 map 15, \*\*\* SEQUENCING  
 DEFINITION  
 AC118710  
 AC118710.6 GI:46849683  
 HTG; HTGS\_PHASE3; HTGS\_FULLTOP; HTGS\_ACTIVEFTN.  
 KEYWORDS Mus musculus (house mouse)  
 SOURCE  
 ORGANISM

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS

1 (bases 1 to 178139)  
 Birren, B., Nusbaum, C. and Lander, E.  
 Mus musculus chromosome 15, clone RP24-121C9  
 Unpublished  
 2 (bases 1 to 178139)  
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,  
 Anderson, S., Barna, N., Baetzten, V., Bloom, T., Boguslavsky, L.,  
 Boukhalter, B., Brown, A., Camarata, J., Campobasso, A., Chang, J.,  
 Chararo, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A.,  
 Cook, A., Cooke, P., DeArrellano, K., Dewar, K., Diaz, J., Dodge, S.,  
 Fero, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,  
 Gande, S., Gird, S., Goyette, M., Graham, L., Grand-Pierre, N.,  
 Hago, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,  
 Kamat, A., Karatas, A., Kells, C., Lachocque, K., Lamazares, R.,  
 Landers, T., Lehoczeky, J., Levine, R., Lindblad-Toh, K., Liu, G.,  
 Maclean, C., MacDonald, P., Major, J., Margolis, N., Matthews, C.,  
 McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Menneus, L.,  
 Mihova, T., Mlenga, V., Murphy, T., Naylor, T., Nguyen, C., Nicol, R.,  
 Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D.,  
 Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,  
 Raymond, C., Retta, R., Rieback, M., Riley, R., Riese, C., Rogov, P.,  
 Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R.,  
 Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,  
 Straus, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,  
 Topham, K., Travers, M., Travis, N., Trifilio, J., Vasilev, H.,  
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,  
 Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS

Direct Submission  
 Submitted (20-APR-2002) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 3 (bases 1 to 178139)  
 Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N.,  
 Anderson, M., Anderson, S., Arachchi, H., Baetzten, V., Bloom, T.,  
 Boukhalter, B., Boukhalter, B., Camarata, J., Chang, J.,  
 Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B.,  
 DeArrellano, K., Diaz, J., Dodge, S., Dooley, K., Dorris, L.,  
 Erikson, J., Fero, S., Ferreira, P., Fitzhugh, W., Gage, D.,  
 Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hago, B.,  
 Hagopian, D., Hago, B., Hall, J., Horton, L., Hulme, W., Iliev, I.,  
 Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T.,  
 Levine, R., Lindblad-Toh, K., Liu, G., Liu, K., Liu, A., Mabbitt, R.,  
 Maclean, C., MacDonald, P., Major, J., Manning, J., Matthews, C.,  
 McCarthy, M., Meldrum, J., Menneus, L., Mihova, T., Mlenga, V.,  
 Murphy, T., Naylor, T., Nguyen, C., Nguyen, C., Nicol, R., Norbu, C.,  
 O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,  
 Phunkhang, P., Pierre, N., Rieback, M., Riese, C., Rogov, P.,  
 Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R.,  
 Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N.,  
 Stojanovic, N., Straus, N., Subramanian, A., Talamas, J., Tesfaye, S.,  
 Theodore, J., Topham, K., Travers, M., Travis, N., Trifilio, J., Vasilev, H.,  
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J.,  
 Zembek, L.,

TITLE  
 JOURNAL  
 COMMENT

Zimmer, A. and Zody, M.  
 Direct Submission  
 Submitted (29-APR-2004) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Apr 29, 2004 this sequence version replaced gi:46367856.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/MIT Center for Genome Research  
 Center code: WIRB  
 Web site: <http://www-seq.wi.mit.edu>  
 Contact: [sequence\\_submissions@road.mit.edu](mailto:sequence_submissions@road.mit.edu)  
 Project Information  
 Center project name: 125058  
 Center clone name: 121\_C\_9

FEATURES  
 source

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 2 contigs. Gaps between the contigs  
 \* are represented as runs of N. The order of the pieces  
 \* is believed to be correct as given, however the sizes  
 \* of the gaps between them are based on estimates that have  
 \* provided by the submitter.  
 \* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.  
 \* 1 66460 66559: contig of 66459 bp in length  
 \* 66560 178139: contig of 111580 bp in length.  
 Location/Qualifiers  
 1. 178139  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10090"  
 /chromosome="15"  
 /map="15"  
 /clone="RP24-121C9"  
 /clone\_lib="RPCT-24 Male Mouse BAC"

## ORIGIN

Query Match 80.9%; Score 17.8; DB 2; Length 178139;  
 Best Local Similarity 90.5%; Pred. No. 2.2e+02;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 ATCTTGACTGCTTAACC 22  
 |||||  
 DB 93272 ATCTTGACTGCTTAAGCC 93292

RESULT 11  
 AP001203 178825 bp DNA linear HTG 29-JUL-2000  
 LOCUS Homo sapiens chromosome 18 clone RP11-845G3 map 18q12, WORKING  
 DEFINITION  
 DRAFT SEQUENCE, 6 unordered pieces.  
 AP001203  
 AP001203.3 GI:9588070  
 HTG; HTGS\_PHASE3; HTGS\_DRAFT.  
 KEYWORDS Homo sapiens (human)  
 SOURCE  
 ORGANISM

REFERENCE  
 AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 178825)  
 Hattori, M., Ishii, K., Toyoda, A., Taylor, T. D., Hong-Seog, P.,  
 Fujiyama, A., Yada, T., Tokoki, Y., Watanabe, H. and Sakaki, Y.

TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS

Homo sapiens 178,825 genomic DNA of 18q12  
 Published Only in Database (2000)  
 2 (bases 1 to 178825)  
 Hattori, M., Ishii, K., Toyoda, A., Taylor, T. D., Hong-Seog, P.,  
 Fujiyama, A., Yada, T., Tokoki, Y., Watanabe, H. and Sakaki, Y.

TITLE  
 JOURNAL

Submitted (18-FEB-2000) Masahita Hattori, The Institute of Physical  
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC),  
 1-15-1 Kitaato, Sagamihara, Kanagawa 228-8555, Japan

## COMMENT

(E-mail: hactori@gsc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/,  
Tel: 81-42-778-9923, Fax: 81-42-778-9924)  
On Jul 31, 2000 this sequence version replaced gi:8117649.

----- Genome Center

Center: RIKEN Genomic Sciences Center (GSC)

Center code: RIKEN

Web site: http://hgp.gsc.riken.go.jp/

Contact: hactori@gsc.riken.go.jp

----- Project Information

Center project name: HumDraFt18

Center clone name: RP11-845G3

----- Summary Statistics

Sequencing vector: PCR products: 100% of reads

Chemistry: Dye-terminator RT-amersham, 100% of reads

Assembly program: Phrap, version 0.990329

Consensus quality: 177242 bases at least Q40

Consensus quality: 177817 bases at least Q30

Consensus quality: 178089 bases at least Q20

Insert size: 178325; sum-of-contigs

Quality coverage: 11.49x in Q20 bases; sum-of-contigs

----- NOTE: This is a 'working draft' sequence. It currently consists of 6 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

```

1       79432 contig of 79432 bp in length
79533   120070 contig of 40538 bp in length
120171  152785 contig of 32615 bp in length
152886  164489 contig of 11604 bp in length
164590  172151 contig of 7562 bp in length
172252  178825 contig of 6574 bp in length.

* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1       79432 contig of 79432 bp in length
79533   120070 contig of 40538 bp in length
120171  152785 contig of 32615 bp in length
152886  164489 contig of 11604 bp in length
164590  172151 contig of 7562 bp in length
172252  178825 contig of 6574 bp in length.

```

## FEATURES

## source

```

1..178825
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="18"
/map="18q12"
/clone="RP11-845G3"

1..79432
/misc_feature
  /note="assembly_fragment"
79533..120070
/misc_feature
  /note="assembly_fragment"
120171..152785
/misc_feature
  /note="assembly_fragment"
152886..164489
/misc_feature
  /note="assembly_fragment"
164590..172151
/misc_feature
  /note="assembly_fragment clone_end:17 vector_side:left"
172252..178825
/misc_feature
  /note="assembly_fragment clone_end:SP6 vector_side:left"

```

## ORIGIN

Query Match 80.9%; Score 17.8; DB 2; Length 178825;  
Best Local Similarity 90.5%; Pred. No. 2, 2e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 ATCTTACTGCTTTAAACC 22

Db 24839 ACTTCTACTGCTTTAAACC 24859

## RESULT 12

## AC101658

LOCUS AC101658 188972 bp DNA linear ROD 30-SEP-2003

DEFINITION Mus musculus chromosome 7, clone RP23-470G18, complete sequence.

## AC101658

AC101658.6 GI:32490721

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

Mus musculus (house mouse)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 188972)

Birken, B., Nusbaum, C. and Lander, E.

## REFERENCE

## JOURNAL

## TITLES

## AUTHORS

## REFERENCE

## JOURNAL

## TITLES

## AUTHORS

## REFERENCE

## JOURNAL

## TITLES

## AUTHORS

## REFERENCE

## JOURNAL

## TITLES

## AUTHORS

## REFERENCE

## JOURNAL

## TITLES

## AUTHORS

## REFERENCE

## JOURNAL

## TITLES

## AUTHORS

## REFERENCE

## JOURNAL

## TITLES

## AUTHORS

## REFERENCE

## JOURNAL

## TITLES

## AUTHORS

## REFERENCE

## JOURNAL

## TITLES

## AUTHORS

## REFERENCE

## JOURNAL

## TITLES

## AUTHORS

## REFERENCE

## JOURNAL

## TITLES

## AUTHORS

## REFERENCE

## JOURNAL

## TITLES

## AUTHORS



**TITLE** Direct Submission  
**JOURNAL** Submitted (05-JUN-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
**REFERENCE** 4 (bases 1 to 188972)  
**AUTHORS** Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N., Anderson, M., Arachchi, H.M., Barna, N., Baetsen, V., Bloom, T., Boguslavsky, L., Boukhalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Cornu, B., Dearellano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hatz, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kelle, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, X., Lui, A., Mabbitt, R., Maclean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Melrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramsamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schnupack, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, C., Stubbs, M., Talamas, J., Teifaye, S., Theodore, J., Topham, K., Travers, M., Vasillejev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X., Wymann, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
**TITLE** Direct Submission  
**JOURNAL** Submitted (10-JUN-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
**REFERENCE** 5 (bases 1 to 188972)  
**AUTHORS** Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N., Anderson, M., Arachchi, H.M., Barna, N., Baetsen, V., Bloom, T., Boguslavsky, L., Boukhalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Cornu, B., Dearellano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hatz, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kelle, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, X., Lui, X., Mabbitt, R., Maclean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Melrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramsamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schnupack, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, C., Stubbs, M., Talamas, J., Teifaye, S., Theodore, J., Topham, K., Travers, M., Vasillejev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X., Wymann, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
**TITLE** Direct Submission  
**JOURNAL** Submitted (30-SEP-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
**COMMENT** On Jul 10, 2003 this sequence version replaced gi:11416098. All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
**TITLE** Direct Submission  
**JOURNAL** Submitted (10-JUN-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
**COMMENT** On Jul 10, 2003 this sequence version replaced gi:11416098. All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
**TITLE** Direct Submission  
**JOURNAL** Submitted (10-JUN-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
**COMMENT** On Jul 10, 2003 this sequence version replaced gi:11416098. All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
**TITLE** Direct Submission  
**JOURNAL** Submitted (10-JUN-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
**COMMENT** On Jul 10, 2003 this sequence version replaced gi:11416098. All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
**TITLE** Direct Submission  
**JOURNAL** Submitted (10-JUN-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
**COMMENT** On Jul 10, 2003 this sequence version replaced gi:11416098. All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
**TITLE** Direct Submission  
**JOURNAL** Submitted (10-JUN-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
**COMMENT** On Jul 10, 2003 this sequence version replaced gi:11416098. All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
**TITLE** Direct Submission  
**JOURNAL** Submitted (10-JUN-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
**COMMENT** On Jul 10, 2003 this sequence version replaced gi:11416098. All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
**TITLE** Direct Submission  
**JOURNAL** Submitted (10-JUN-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
**COMMENT** On Jul 10, 2003 this sequence version replaced gi:11416098. All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
**TITLE** Direct Submission  
**JOURNAL** Submitted (10-JUN-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
**COMMENT** On Jul 10, 2003 this sequence version replaced gi:11416098. All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
**TITLE** Direct Submission  
**JOURNAL** Submitted (10-JUN-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
**COMMENT** On Jul 10, 2003 this sequence version replaced gi:11416098. All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
**TITLE** Direct Submission  
**JOURNAL** Submitted (10-JUN-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
**COMMENT** On Jul 10, 2003 this sequence version replaced gi:11416098. All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
**TITLE** Direct Submission  
**JOURNAL** Submitted (10-JUN-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
**COMMENT** On Jul 10, 2003 this sequence version replaced gi:11416098. All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
**TITLE** Direct Submission  
**JOURNAL** Submitted (10-JUN-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
**COMMENT** On Jul 10, 2003 this sequence version replaced gi:11416098. All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
**TITLE** Direct Submission  
**JOURNAL** Submitted (10-JUN-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
**COMMENT** On Jul 10, 2003 this sequence version replaced gi:11416098. All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
**TITLE** Direct Submission  
**JOURNAL** Submitted (10-JUN-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
**COMMENT** On Jul 10, 2003 this sequence version replaced gi:11416098. All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
**TITLE** Direct Submission  
**JOURNAL** Submitted (10-JUN-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
**COMMENT** On Jul 10, 2003 this sequence version replaced gi:11416098. All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
**TITLE** Direct Submission  
**JOURNAL** Submitted (10-JUN-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
**COMMENT** On Jul 10, 2003 this sequence version replaced gi:11416098. All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
**TITLE** Direct Submission  
**JOURNAL** Submitted (10-JUN-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
**COMMENT** On Jul 10, 2003 this sequence version replaced gi:11416098. All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
**TITLE** Direct Submission  
**JOURNAL** Submitted (10-JUN-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
**COMMENT** On Jul 10, 2003 this sequence version replaced gi:11416098. All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
**TITLE** Direct Submission

misc_feature	1. .569	/note="wgs_end_extension
	clone_end:SP6"	
repeat_region	complement(109. .148)	
	/rpt_family="MIR"	
misc_feature	570. .575	/note="clone boundary
	clone_end:SP6	
	site:ECORI"	
repeat_region	1189. .1209	/rpt_family="AT_rich"
	complement(2030. .2229)	
repeat_region	/rpt_family="B3A"	
	complement(2713. .2812)	
repeat_region	/rpt_family="MERSB"	
	complement(1596. .3709)	
repeat_region	/rpt_family="MERSA"	
	3864. .3988	/rpt_family="B1F"
repeat_region	3983. .4116	/rpt_family="B3"
	4175. .4275	
repeat_region	/rpt_family="(TA)n"	
	4275. .4301	
repeat_region	/rpt_family="(GA)n"	
	4315. .4387	
repeat_region	/rpt_family="(GGAA)n"	
	4771. .4868	
repeat_region	/rpt_family="(GA-rich"	
	complement(5908. .5967)	
repeat_region	/rpt_family="MIR"	
	6172. .6459	
repeat_region	/rpt_family="B4A"	
	6471. .6516	
repeat_region	/rpt_family="(CAA)n"	
	6522. .6633	
repeat_region	/rpt_family="NSINE1"	
	9347. .9913	
repeat_region	/rpt_family="Lx2"	
	complement(9976. .10192)	
repeat_region	/rpt_family="L2"	
	10379. .10483	
repeat_region	/rpt_family="MNER15"	
	10563. .10748	
repeat_region	/rpt_family="MNER15"	
	complement(11784. .12143)	
repeat_region	/rpt_family="MSTD"	
	12239. .12854	
repeat_region	/rpt_family="MNER19B"	
	complement(12857. .13275)	
repeat_region	/rpt_family="MTC"	
	13370. .13487	
repeat_region	/rpt_family="MHRV16"	
	13515. .13667	
repeat_region	/rpt_family="B3"	
	14790. .15191	
repeat_region	/rpt_family="MTC"	
	complement(15464. .15740)	
repeat_region	/rpt_family="MUTLE"	
	complement(15790. .15844)	
repeat_region	/note="single clone coverage"	
	15803. .16116	
repeat_region	/rpt_family="MTE"	
	complement(16117. .16335)	
repeat_region	/rpt_family="URRLB"	
	16336. .16427	
repeat_region	/rpt_family="MTE"	

Query Match 80.9% Score 17.8; DB 10; Length 188972;  
Best Local Similarity 90.5% Pred.No.2.1e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;



```

Db      179704 ATCTGTACTGCTTTAACCC 179724

RESULT 13
BX470182
LOCUS    BX470182
DEFINITION Danto rerio clone CH211-195P4, WORKING DRAFT SEQUENCE, 3 unordered
pieces.
ACCESSION BX470182.7 GI:508313349
VERSION   HTG; HTGS PHASE1; HTGS ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS  Danto rerio (zebrafish)
SOURCE    Danto rerio
ORGANISM  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danto.
1 (bases 1 to 208244)
Mashreqhi-Mohammadi, M.
Direct Submission
Submitted (28-JUL-2004) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Jul 29, 2004 this sequence version replaced gi:49614046.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: zfish-help@sanger.ac.uk
----- Project Information
Center project name: zc195P4
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 207900 bases at least Q40
Consensus quality: 207989 bases at least Q30
Consensus quality: 208016 bases at least Q20
Insert size: 208044; sum-of-contigs
Insert size: 203672; 7.6% error; agarose-fp
Quality coverage: 10.93x in Q20 bases; sum-of-contigs Quality
coverage: 11.26x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
1. 8185: contig of 8185 bp in length
*
* 8186 8285: gap of 100 bp
*
* 8286 146816: contig of 138531 bp in length
*
* 146817 146916: gap of 100 bp
*
* 146917 208244: contig of 61328 bp in length.
Location/Qualifiers
1. 208244
/organism="Danto rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="CH211-195P4"
/clone_1lb="CHORI-211"
1. 8185
/notes="assembly fragment: 00104
fragment chain: 1
clone end: SP6
vector side: left"
8286..146816
/notes="assembly fragment: 03435
fragment chain: 1"
146917..208244
/notes="assembly fragment: 00192
fragment chain: 1"
FEATURES
. source
m18c_feature
m18c_feature
m18c_feature
m18c_feature

```

```

ORIGIN
Query Match      80.5%; Score 17.8; DB 2; Length 208244;
Best Local Similarity 90.5%; Pred. No. 2.1e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Cy      2 ATCTGTACTGCTTTAACCC 22
Db      170912 ATCTGTACTGCTTTAACCC 170932

RESULT 14
AC110680
LOCUS    AC110680
DEFINITION Rattus norvegicus clone CH230-250H8, WORKING DRAFT SEQUENCE.
ACCESSION AC110680.5 GI:25139138
VERSION   HTG; HTGS PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS  Rattus norvegicus (Norway rat)
SOURCE    Rattus norvegicus
ORGANISM  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 213071)
Muzny, D., Marie, Metzker, M., Lee, Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, B., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dedrich, D.,
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Diya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Frazer, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Gebreyes, E., Geer, K., Gill, R., Grady, M., Guerra, M., Guevara, W.,
Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogue, M.,
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, J., Kovar, C.,
Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorenz, L., Louised, H., Lozano, R., Lu, X., Ma, J.,
Maheshwari, M., Mahindaratne, M., Mahmoud, M., Mallory, K., Mangum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mawhney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Muniasa, M., Murphy, M., Nair, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
Nwoko, E., O., Okwumodu, G., Olarinmashogun, A., Pal, S., Parks, K.,
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L.,
Punzo, M., Quiroz, J., Rachlin, R., Reeves, K., Regier, M.A., Reigh, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
River, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,
Sanders, M., Savery, G., Scherer, S., Scott, G., Shattman, S., Shen, H.,
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smales, D.,
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Soza, J.,
Steinle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Tyanji, K.,
Valas, R., Vera, V., Villaseca, D., Waldron, L., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
Williams, G., Willson, R., Wleczek, R., Wooden, H., Wortley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhou, S., Dunn, D., von
Niederhausen, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,
Weinstock, G. and Gibbs, R.A.

```

```

TITLE      Direct Submission
JOURNAL    Unpublished
REFERENCE   2 (bases 1 to 213071)
AUTHORS    Worley,K.C.
TITLE      Direct Submission
JOURNAL    Submitted (15-FEB-2002) Human Genome Sequencing Center, Department
REFERENCE   of Molecular and Human Genetics, Baylor College of Medicine, One
AUTHORS     Baylor Plaza, Houston, TX 77030, USA
TITLE       3 (bases 1 to 213071)
JOURNAL     Rat Genome Sequencing Consortium.
REFERENCE   Direct Submission
AUTHORS     Submitted (20-NOV-2002) Human Genome Sequencing Center, Department
TITLE       of Molecular and Human Genetics, Baylor College of Medicine, One
JOURNAL     Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 20, 2002 this sequence version replaced gi:21335098.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.

            ----- Genome Center
            Center: Baylor College of Medicine
            Center code: BCM
            Web site: http://www.hgsc.bcm.tmc.edu/
            Contact: hgsc-help@bcm.tmc.edu
            ----- Project Information
            Center project name: GQVF
            Center clone name: CH230-25048
            ----- Summary Statistics
            Assembly program: Phrap; version 0.990329
            Consensus quality: 202669 bases at least Q40
            Consensus quality: 204833 bases at least Q30
            Consensus quality: 206266 bases at least Q20
            Estimated insert size: 207285; sum-of-contigs estimation
            Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

            -----
            * NOTE: Estimated insert size may differ from sequence length
            * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html) .
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            1 213071: contig of 213071 bp in length.
               Location/Qualifiers
                 1. .213071
                    /organism="Rattus norvegicus"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:10116"
                    /clone="CH230-250H8"
                 1. .2703
                    /note="wgs_end_extension
clone_end:T7"
84612_ .84647
/note="Clone_boundary
clone_end:T7
site:
end_sequence:BZ219211"
complement(206636..207330)
/note="clone boundary
clone_end:Sp6
site:

```

```

misc_feature          end_sequence:82219213"
                       208785. .209961
                       /note="wgs_end_extension
misc_feature          clone_end:5p6
                       211965. .213071
                       /note="wgs_end_extension
                       clone_end:5p6"
ORIGIN
Query Match          80.9%; Score 17.8; DB 2; Length 213071;
Best Local Similarity 90.5%; Pred. No. 2.1e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY          1 GATCTTGACTGCTCTTAACC 21
             |||||
             |||||
DB          141740 GATCTTGACTGCTCTCAACC 141760

RESULT 15
AC087150          223538 bp      DNA      linear      HTG 09-DEC-2000
LOCUS          AC087150/c
DEFINITION      Mus musculus clone RP23-312B8, WORKING DRAFT SEQUENCE, 20 unordered
                 pieces.
ACCESSION      AC087150
VERSION        AC087150.1 GI:11610876
KEYWORDS       HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE         Mus musculus (house mouse)
ORGANISM       Mus musculus
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
REFERENCE      1 (bases 1 to 223538)
AUTHORS        DOE Joint Genome Institute.
TITLE          Sequencing of Mouse
JOURNAL        2 (bases 1 to 223538)
REFERENCE      DOE Joint Genome Institute.
AUTHORS        Direct Submission
TITLE          Submitted (09-DEC-2000) Production Sequencing Facility, DOE Joint
JOURNAL        Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT        -----Genome Center
                Center: Joint Genome Institute
                Center Code: JGI
                Web site: http://www.jgi.doe.gov
                -----
                Project Information
                Center Project Name: 1856712
                Center clone name: RPCI-23_312B8
                -----
                Summary Statistics
                Consensus quality: 207168 bases at least Q40
                Consensus quality: 211223 bases at least Q30
                Consensus quality: 213792 bases at least Q20
                Estimated insert size: 210000; agarose-fp estimation
                Estimated insert size: 221638; sum-of-contigs estimation
                Quality coverage: 10.16 in Q20 bases; agarose-fp estimation
                Quality coverage: 9.63 in Q20 bases; sum-of-contigs estimation.
                * NOTE: This is a 'working draft' sequence. It currently
                * consists of 20 contigs. The true order of the pieces
                * is not known and their order in this sequence record is
                * arbitrary. Gaps between the contigs are represented as
                * runs of N, but the exact sizes of the gaps are unknown.
                * This record will be updated with the finished sequence
                * as soon as it is available and the accession number will
                * be preserved.
                -----
                1
                1039: contig of 1038 bp in length
                1138: gap of unknown length
                1139: contig of 1068 bp in length
                2207: gap of unknown length
                2307: contig of 1218 bp in length
                3524: gap of unknown length
                3525: contig of 1573 bp in length
                5198: gap of unknown length
                5297: contig of 1094 bp in length
                6391: contig of 1094 bp in length

```

```

* 6392 6491: gap of unknown length
* 6492 8371: contig of 1880 bp in length
* 8372 8372 8471: gap of unknown length
* 8472 9651: contig of 1180 bp in length
* 9652 9751: gap of unknown length
* 9752 11247: contig of 1496 bp in length
* 11348 11347: gap of unknown length
* 11348 12711: contig of 1364 bp in length
* 12712 12811: gap of unknown length
* 12812 18292: contig of 5481 bp in length
* 18293 18392: gap of unknown length
* 18393 27516: contig of 9124 bp in length
* 27517 27616: gap of unknown length
* 27617 37488: contig of 9872 bp in length
* 37489 37588: gap of unknown length
* 37589 52023: contig of 14435 bp in length
* 52024 52123: gap of unknown length
* 52124 69050: contig of 16927 bp in length
* 69051 69150: gap of unknown length
* 69151 91246: contig of 22096 bp in length
* 91247 91346: gap of unknown length
* 91347 106178: contig of 14832 bp in length
* 106179 106278: gap of unknown length
* 106279 131304: contig of 25026 bp in length
* 131305 131404: gap of unknown length
* 131405 153810: contig of 22406 bp in length
* 153811 153910: gap of unknown length
* 153911 183927: contig of 30017 bp in length
* 183928 184027: gap of unknown length
* 184028 223538: contig of 39511 bp in length.

```

## FEATURES

## Source

```

1..223538
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone="RP23-312B8"
/location="RP23-312B8"
/clone_lib="RP23-312B8"

```

## ORIGIN

```

Query Match      80.9%; Score 17.8; DB 2; Length 223538;
Best Local Similarity 90.5%; Pred. No. 2.1e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 GATCTGTACTGCTTAACC 21
DB 161710 GATCTGTGACTGCTTAAGACC 161690

```

Search completed: February 8, 2005, 16:36:51  
 Job time : 133.665 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using bw model

Run on: February 8, 2005, 11:25:10 : Search time 29.1839 Seconds  
(without alignments)  
3957.220 Million cell updates/sec

Title: US-10-754-437-5

Perfect score: 22  
Sequence: 1 gacctgctactgtcttaaaccc 22

Scoring table: IDENTITY\_NTC  
Gapop 10.0, Gapext 1.0

Searched: 413486 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_23sep04:\*  
1: geneseqn1980s:\*  
2: geneseqn1990s:\*  
3: geneseqn2000s:\*  
4: geneseqn2001as:\*  
5: geneseqn2001bs:\*  
6: geneseqn2002as:\*  
7: geneseqn2002bs:\*  
8: geneseqn2003as:\*  
9: geneseqn2003bs:\*  
10: geneseqn2003cs:\*  
11: geneseqn2003ds:\*  
12: geneseqn2004s:\*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	17.2	78.2	363	2	AAX41422 Human sec
C 2	17.2	78.2	454	2	ACH22743 Human adu
C 3	17.2	78.2	457	3	AAC55151 Arabidops
C 4	17.2	78.2	470	6	AB193436 Arabidops
C 5	17.2	78.2	636	3	AAC55036 Arabidops
C 6	17.2	78.2	641	3	AAC53341 Arabidops
C 7	17.2	78.2	101569	12	ADP13274 Renal cel
C 8	17.2	78.2	259202	12	ADQ18492 Human sof
C 9	16.8	76.4	567	6	ABN78623 Human byn
C 10	16.8	76.4	637	6	ABN61229 Human can
C 11	16.8	76.4	1452	10	ACC60893 Gene begu
C 12	16.8	76.4	1452	10	ADK62303 Disease t
C 13	16.4	74.5	356	3	AAC29428 Human sec
C 14	16.4	74.5	508	4	ABA60565 Human fce
C 15	16.4	74.5	508	4	AA140454 Probe #91
C 16	16.4	74.5	508	4	AAX43735 Human bon
C 17	16.4	74.5	508	4	AAX08847 Human bra
C 18	16.4	74.5	139257	10	ADC89520 Human COR
C 19	16.2	73.6	199	10	ABH84623 Corn ear-
C 20	16.2	73.6	339	4	AAX53723 Murine tr
C 21	16.2	73.6	473	8	ABZ55713 Aspergill

22	16.2	73.6	499	9	AC119831	AC119831 DNA clone
23	16.2	73.6	510	9	AC119828	AC119828 DNA clone
24	16.2	73.6	530	9	AC119827	AC119827 DNA clone
25	16.2	73.6	534	9	AC119830	AC119830 DNA clone
26	16.2	73.6	536	9	AC119835	AC119835 DNA clone
27	16.2	73.6	557	3	AAC53403	AAC53403 Arabidops
28	16.2	73.6	565	9	AC119840	AC119840 DNA clone
29	16.2	73.6	578	9	AC119855	AC119855 DNA clone
30	16.2	73.6	588	2	ADR02268	ADR02268 A. gossyp
31	16.2	73.6	590	9	AC119847	AC119847 DNA clone
32	16.2	73.6	591	9	AC119858	AC119858 DNA clone
33	16.2	73.6	595	9	AC119839	AC119839 DNA clone
34	16.2	73.6	609	9	AC119854	AC119854 DNA clone
35	16.2	73.6	618	9	AC119838	AC119838 DNA clone
36	16.2	73.6	659	9	AC119864	AC119864 DNA clone
37	16.2	73.6	661	9	AC119860	AC119860 DNA clone
38	16.2	73.6	662	9	AC119837	AC119837 DNA clone
39	16.2	73.6	667	9	AC119853	AC119853 DNA clone
40	16.2	73.6	675	9	AC119853	AC119853 DNA clone
41	16.2	73.6	682	9	AC119877	AC119877 DNA clone
42	16.2	73.6	689	9	AC119874	AC119874 DNA clone
43	16.2	73.6	690	9	AC119867	AC119867 DNA clone
44	16.2	73.6	691	3	AAC54484	AAC54484 Arabidops
45	16.2	73.6	694	9	AC119862	AC119862 DNA clone

## ALIGNMENTS

RESULT 1	AAX41422/c
ID	AAX41422 standard; cDNA; 363 BP.
AC	AAX41422;
XX	
DT	22-JUN-1999 (first entry)
XX	
DE	Human secreted protein 5' EST SEQ ID NO: 81 from WO 9906553.
XX	
KW	Human; secreted protein; EST; expressed sequence tag; diagnosis;
XX	
KW	forensic; gene therapy; chromosome mapping; signal peptide;
XX	
KW	upstream regulatory sequence; cytokine activity; cell proliferation;
XX	
KW	differentiation; haematopoiesis regulation; tissue growth regulation;
XX	
KW	reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
XX	
KW	thrombolytic; antiinflammatory; tumour inhibition; antitumour; de.
XX	
OS	Homo sapiens.
XX	
PN	WO9906553-A2.
XX	
PD	11-FEB-1999.
XX	
PF	31-JUL-1998; 98WO-1B001237.
XX	
PR	01-AUG-1997; 97US-00905051.
XX	
PA	(GEST ) GENSET.
XX	
PI	Dumas Milne Edwards J, Duclert A, Lacroix B;
XX	
DR	WPI; 1999-153783/13.
XX	
DR	P-PsDB; AAY12564.
XX	
PT	New nucleic acids encoding human secreted proteins - obtained from cDNA
XX	
PT	libraries derived from umbilical cord, lymph ganglia, lymphocytes and
XX	
PS	placental tissue.
XX	
PS	Claim 1, Page 200; 41pp; English.
XX	
CC	AAX41379 to AAX41526 represent 5' expressed sequence tags (ESTs) for
XX	
CC	human secreted proteins, and encode the proteins given in AAY12521 to
XX	
CC	AAY12668, respectively. The proteins given represent the signal peptide
XX	
CC	and an N-terminal fragment of a secreted protein. The nucleic acid

CC sequences can be used for producing secreted human gene products. They  
CC can also be used to develop products for diagnosis and therapy. The  
CC proteins obtained may have cytokine activity, cell  
CC proliferation/differentiation activity, haematopoiesis regulating  
CC activity, tissue growth regulating activity, reproductive hormone  
CC regulating activity, chemotactic/chemokinetic activity, haemostatic and  
CC thrombolytic activity, receptor/ligand activity, antiinflammatory  
CC activity, tumour inhibition activity or other activities. The products  
CC can be used in forensic, gene therapy and chromosome mapping procedures.  
CC The sequences can also be used for obtaining corresponding promoter  
CC sequences. The nucleic acids encoding the signal peptide can be used for  
CC directing extracellular secretion of a polypeptide or the insertion of a  
CC polypeptide into a membrane, or importing a polypeptide into a cell  
CC  
SQ Sequence 363 BP; 114 A; 79 C; 81 G; 89 T; 0 U; 0 Other;

Query Match 78.2%; Score 17.2; DB 2; Length 363;  
Best Local Similarity 86.4%; Pred. No. 1.7e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 GATCTGTACTGCTTAAACC 22  
DB 169 GATCTGTCTGCTTAACTC 148

RESULT 2  
ACH22743/c  
ID ACH22743 standard; cDNA; 454 BP.  
AC ACH22743;  
XX  
DT 13-OCT-2003 (first entry)  
XX  
DE Human adult ovary cDNA #1123.  
XX  
KW Human; ss; sequencing by hybridisation; SSH; expressed sequence tag; EST;  
XX genome mapping; biodiversity; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PN US2003073623-A1.  
XX  
PD 17-APR-2003.  
XX  
PF 30-JUL-2001; 2001US-00918995.  
XX  
PR 30-JUL-2001; 2001US-00918995.  
XX  
PA (DRMA/) DRMANAC R T.  
PA (LABA/) LABAT I.  
PA (STAC/) STACHE-CRAIN B.  
PA (DICK/) DICKSON M C.  
PA (JONE/) JONES L W.  
PI Dramac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;  
XX WPI; 2003-615964/58.  
XX  
DR  
PT New polynucleotide sequences obtained from various cDNA libraries, useful  
PT as hybridization probes, as oligomers for PCR, for chromosome and gene  
PT mapping, in the recombinant production of protein, or in generating  
PT antisense DNA or RNA.  
XX  
PS Claim 1; SEQ ID NO 9955; 44pp; English.  
XX  
CC The invention relates to an isolated polynucleotide comprising any one of  
CC 38043 cDNA sequences, appearing as ACH22743-ACH50831, whose sequence was  
CC determined by the technique of SSH (sequencing by hybridisation). Also  
CC included is a purified polypeptide comprising a sequence corresponding to  
CC a reading frame of the novel polynucleotide. The nucleic acid sequences  
CC are useful in diagnostics as expressed sequence tags (EST) for  
CC identifying expressed genes or for physical mapping of the human genome,  
CC in forensics, in assessing biodiversity, or in identifying mutations

CC responsible for genetic disorders and other traits. The nucleotide  
CC sequences are also useful as hybridisation probes, as oligomers for PCR,  
CC for chromosome and gene mapping, in the recombinant production of  
CC protein, or in generating antisense DNA or RNA. The purified polypeptide  
CC is useful for generating antibodies specific for it. The present sequence  
CC is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data  
CC for this patent did not form part of the printed specification, but was  
CC obtained in electronic format directly from USPTO at  
CC [seqdata.uspto.gov/sequence.html?DocID=20030073623](http://seqdata.uspto.gov/sequence.html?DocID=20030073623)  
CC  
SQ Sequence 454 BP; 149 A; 89 C; 103 G; 113 T; 0 U; 0 Other;

Query Match 78.2%; Score 17.2; DB 9; Length 454;  
Best Local Similarity 86.4%; Pred. No. 1.7e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 GATCTGTACTGCTTAAACC 22  
DB 140 GATCTGTCTGCTTAACTC 119

RESULT 3  
AAC5151/c  
ID AAC5151 standard; DNA; 457 BP.  
AC AAC5151;  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 80200.  
XX  
KW Hybridisation assay; genetic mapping; gene expression control;  
XX protein identification; signal transduction pathway; metabolic pathway;  
XX promoter; termination sequence; ss.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-00301439.  
XX  
PR 25-FEB-1999; 99US-0121825P.  
XX  
PR 09-MAR-1999; 99US-0123180P.  
XX  
PR 23-MAR-1999; 99US-0123548P.  
XX  
PR 25-MAR-1999; 99US-0125788P.  
XX  
PR 29-MAR-1999; 99US-0126264P.  
XX  
PR 01-APR-1999; 99US-0127462P.  
XX  
PR 06-APR-1999; 99US-0128234P.  
XX  
PR 08-APR-1999; 99US-0128714P.  
XX  
PR 16-APR-1999; 99US-0128845P.  
XX  
PR 19-APR-1999; 99US-0130077P.  
XX  
PR 21-APR-1999; 99US-0130449P.  
XX  
PR 23-APR-1999; 99US-0130510P.  
XX  
PR 28-APR-1999; 99US-0131449P.  
XX  
PR 30-APR-1999; 99US-0132048P.  
XX  
PR 30-APR-1999; 99US-0132407P.  
XX  
PR 04-MAY-1999; 99US-0132484P.  
XX  
PR 05-MAY-1999; 99US-0132485P.  
XX  
PR 06-MAY-1999; 99US-0132486P.  
XX  
PR 06-MAY-1999; 99US-0132487P.  
XX  
PR 07-MAY-1999; 99US-0132863P.  
XX  
PR 11-MAY-1999; 99US-0134256P.  
XX  
PR 14-MAY-1999; 99US-0134218P.  
XX  
PR 14-MAY-1999; 99US-0134221P.  
XX  
PR 14-MAY-1999; 99US-0134370P.  
XX  
PR 18-MAY-1999; 99US-0134768P.  
XX  
PR 19-MAY-1999; 99US-0134941P.  
XX  
PR 20-MAY-1999; 99US-0135124P.

```
PR 21-MAY-1999; 99US-0135353P-
PR 24-MAY-1999; 99US-0135629P-
PR 25-MAY-1999; 99US-0136021P-
PR 27-MAY-1999; 99US-0136392P-
PR 28-MAY-1999; 99US-0136782P-
PR 01-JUN-1999; 99US-0137222P-
PR 03-JUN-1999; 99US-0137528P-
PR 04-JUN-1999; 99US-0137502P-
PR 07-JUN-1999; 99US-0137724P-
PR 08-JUN-1999; 99US-0138094P-
PR 10-JUN-1999; 99US-0138540P-
PR 10-JUN-1999; 99US-0138847P-
PR 14-JUN-1999; 99US-0139119P-
PR 16-JUN-1999; 99US-0139452P-
PR 16-JUN-1999; 99US-0139453P-
PR 17-JUN-1999; 99US-0139492P-
PR 18-JUN-1999; 99US-0139454P-
PR 18-JUN-1999; 99US-0139455P-
PR 18-JUN-1999; 99US-0139456P-
PR 18-JUN-1999; 99US-0139457P-
PR 18-JUN-1999; 99US-0139458P-
PR 18-JUN-1999; 99US-0139459P-
PR 18-JUN-1999; 99US-0139460P-
PR 18-JUN-1999; 99US-0139461P-
PR 18-JUN-1999; 99US-0139462P-
PR 18-JUN-1999; 99US-0139463P-
PR 18-JUN-1999; 99US-0139750P-
PR 18-JUN-1999; 99US-0139763P-
PR 21-JUN-1999; 99US-0139817P-
PR 22-JUN-1999; 99US-0139899P-
PR 23-JUN-1999; 99US-0140353P-
PR 23-JUN-1999; 99US-0140354P-
PR 24-JUN-1999; 99US-0140695P-
PR 28-JUN-1999; 99US-0140823P-
PR 29-JUN-1999; 99US-0140919P-
PR 30-JUN-1999; 99US-0141287P-
PR 01-JUL-1999; 99US-0141842P-
PR 01-JUL-1999; 99US-0142154P-
PR 02-JUL-1999; 99US-0142055P-
PR 06-JUL-1999; 99US-0142390P-
PR 08-JUL-1999; 99US-0142803P-
PR 09-JUL-1999; 99US-0142920P-
PR 12-JUL-1999; 99US-0142977P-
PR 13-JUL-1999; 99US-0143542P-
PR 14-JUL-1999; 99US-0143624P-
PR 15-JUL-1999; 99US-0144005P-
PR 16-JUL-1999; 99US-0144085P-
PR 19-JUL-1999; 99US-0144325P-
PR 19-JUL-1999; 99US-0144331P-
PR 19-JUL-1999; 99US-0144332P-
PR 19-JUL-1999; 99US-0144333P-
PR 19-JUL-1999; 99US-0144334P-
PR 19-JUL-1999; 99US-0144335P-
PR 20-JUL-1999; 99US-0144352P-
PR 20-JUL-1999; 99US-0144632P-
PR 20-JUL-1999; 99US-0144884P-
PR 21-JUL-1999; 99US-0144814P-
PR 21-JUL-1999; 99US-0145086P-
PR 21-JUL-1999; 99US-0145088P-
PR 22-JUL-1999; 99US-0145085P-
PR 22-JUL-1999; 99US-0145087P-
PR 22-JUL-1999; 99US-0145089P-
PR 22-JUL-1999; 99US-0145192P-
PR 23-JUL-1999; 99US-0145145P-
PR 23-JUL-1999; 99US-0145218P-
PR 23-JUL-1999; 99US-0145224P-
PR 26-JUL-1999; 99US-0145276P-
PR 27-JUL-1999; 99US-0145813P-
PR 27-JUL-1999; 99US-0145918P-
PR 27-JUL-1999; 99US-0145919P-
PR 28-JUL-1999; 99US-0145951P-
PR 02-AUG-1999; 99US-0146386P-
PR 02-AUG-1999; 99US-0146388P-
PR 02-AUG-1999; 99US-0147038P-
PR 03-AUG-1999; 99US-0147204P-
PR 04-AUG-1999; 99US-0147302P-
PR 04-AUG-1999; 99US-0147329P-
PR 05-AUG-1999; 99US-0147260P-
PR 06-AUG-1999; 99US-0147303P-
PR 06-AUG-1999; 99US-0147416P-
PR 09-AUG-1999; 99US-0147493P-
PR 09-AUG-1999; 99US-0147935P-
PR 10-AUG-1999; 99US-0148171P-
PR 11-AUG-1999; 99US-0148319P-
PR 12-AUG-1999; 99US-0148341P-
PR 13-AUG-1999; 99US-0148865P-
PR 13-AUG-1999; 99US-0148884P-
PR 16-AUG-1999; 99US-0149368P-
PR 17-AUG-1999; 99US-0149175P-
PR 18-AUG-1999; 99US-0149426P-
PR 20-AUG-1999; 99US-0149722P-
PR 20-AUG-1999; 99US-0149723P-
PR 20-AUG-1999; 99US-0149929P-
PR 23-AUG-1999; 99US-0149902P-
PR 25-AUG-1999; 99US-0149930P-
PR 25-AUG-1999; 99US-0150566P-
PR 26-AUG-1999; 99US-0150884P-
PR 27-AUG-1999; 99US-0151065P-
PR 27-AUG-1999; 99US-0151066P-
PR 27-AUG-1999; 99US-0151080P-
PR 30-AUG-1999; 99US-0151030P-
PR 31-AUG-1999; 99US-0151438P-
PR 01-SEP-1999; 99US-0151930P-
PR 07-SEP-1999; 99US-0152263P-
PR 10-SEP-1999; 99US-0153070P-
PR 13-SEP-1999; 99US-0153758P-
PR 15-SEP-1999; 99US-0154018P-
PR 16-SEP-1999; 99US-0154039P-
PR 20-SEP-1999; 99US-0154799P-
PR 22-SEP-1999; 99US-0155139P-
PR 23-SEP-1999; 99US-0155486P-
PR 28-SEP-1999; 99US-0155659P-
PR 29-SEP-1999; 99US-0156458P-
PR 29-SEP-1999; 99US-0156596P-
PR 04-OCT-1999; 99US-0157117P-
PR 05-OCT-1999; 99US-0157753P-
PR 06-OCT-1999; 99US-0157865P-
PR 07-OCT-1999; 99US-0158029P-
PR 08-OCT-1999; 99US-0158232P-
PR 12-OCT-1999; 99US-0158369P-
PR 13-OCT-1999; 99US-0159293P-
PR 13-OCT-1999; 99US-0159294P-
PR 13-OCT-1999; 99US-0159295P-
PR 14-OCT-1999; 99US-0159329P-
PR 14-OCT-1999; 99US-0159330P-
PR 14-OCT-1999; 99US-0159331P-
PR 14-OCT-1999; 99US-0159637P-
PR 14-OCT-1999; 99US-0159638P-
PR 18-OCT-1999; 99US-0159584P-
PR 21-OCT-1999; 99US-0160741P-
PR 21-OCT-1999; 99US-0160767P-
PR 21-OCT-1999; 99US-0160768P-
PR 21-OCT-1999; 99US-0160770P-
PR 21-OCT-1999; 99US-0160814P-
PR 21-OCT-1999; 99US-0160815P-
PR 22-OCT-1999; 99US-0160980P-
PR 22-OCT-1999; 99US-0160981P-
PR 22-OCT-1999; 99US-0160989P-
PR 25-OCT-1999; 99US-0161404P-
PR 25-OCT-1999; 99US-0161405P-
PR 25-OCT-1999; 99US-0161406P-
PR 26-OCT-1999; 99US-0161359P-
PR 26-OCT-1999; 99US-0161360P-
PR 26-OCT-1999; 99US-0161361P-
```

```
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match      78.2%; Score 17.2; DB 3; Length 457;
Best Local Similarity 86.4%; Pred. No. 1.7e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GATCTTGACGTGCTTAACCC 22
DB 189 GATCTTGTTGTCATTAACCC 168

RESULT 4
ABL93436
ID ABL93436 standard; cDNA; 470 BP.
AC ABL93436;
XX
XX 10-JUN-2002 (first entry)
XX
DE Arabidopsis thaliana nucleic acid sequence Ref:2027201 SEQ ID NO:201.
XX
XX Arabidopsis thaliana; insecticide; fungicide; plant; mapping; diagnosis;
KM genetic modification; gene; ss.
XX
OS Arabidopsis thaliana.
XX
XX US2002023280-A1.
XX
XX 21-FEB-2002.
XX
XX 26-JAN-2001; 2001US-00770444.
XX
XX 27-JAN-2000; 2000US-0178502P.
XX
XX (GORL/) GORLACH J.
XX (ANY/) AN Y.
XX (HAM/) HAMILTON C M.
XX (PRIC/) PRICE J L.
XX (RAIN/) RAINES T M.
XX (YUY/) YU Y.
XX (RAME/) RAMEAKA J G.
XX (PAGE/) PAGE A.
XX (MATH/) MATHEW A V.
XX (LEDF/) LEDFORD B L.
XX (WOES/) WOESSNER J P.
XX (HAAS/) HAAS W D.
XX (GARC/) GARCIA C A.
XX (KRIC/) KRICKER M.
XX (SLAT/) SLATER T.
XX (DAVI/) DAVIS K R.
XX (ALLE/) ALLEN K.
XX (HOFF/) HOFFMAN N.
XX (HURB/) HURBAN P.
XX
XX Gorlach J, An Y, Hamilton CM, Price JT, Raines TM, Yu Y;
XX Rameaka JG, Page A, Mathew AV, Ledford BU, Woessner JP, Haas WD;
XX Garcia CA, Kricker M, Slater T, Davis KR, Allen K, Hoffman N;
XX Hurban P;
XX
XX WPI; 2002-267486/31.
XX
XX New Arabidopsis thaliana nucleic acid, for identifying homologous genes,
XX producing compositions that modulate the expression or function of its
XX encoded protein, and mapping functional regions of a protein.
XX
XX Claim 1; SEQ ID NO 201; 44pp; English.
XX
XX The present invention describes an Arabidopsis thaliana nucleic acid (I)
XX comprising a sequence capable of hybridizing under stringent conditions
XX to a sequence (SI) selected from any one of the 999 sequences given in
```

```
CC ABL93236 to ABL94234. (I) have insecticide and fungicide activities, and
CC they can be used as protein expression modulators. (I) can be used in
CC identifying homologous or related genes, in producing compositions that
CC modulate the expression or function of their encoded proteins, mapping
CC functional regions of the protein, and in studying associated
CC physiological pathways. (I) can also be used: (1) for the genetic
CC manipulation of cells, particularly plant cells; (2) in screening assays
CC of various plant strains to determine the strains that are best capable
CC of withstanding a particular disease or environmental stress; (3) for
CC enhancing or inhibiting production of a biosynthetic product in a plant;
CC (4) as probes in mapping and in diagnosis, in genetic modification and
CC for screening purposes, to generate additional copies of the nucleic
CC acids, to generate ribozymes or antisense oligonucleotides, and as single
CC stranded DNA probes or as triple-strand forming oligonucleotides; and (5)
CC for generating genetically modified transgenic organisms. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC USPTO web site
XX
SQ Sequence 470 BP; 116 A; 107 C; 95 G; 152 T; 0 U; 0 Other;
XX
XX Query Match      78.2%; Score 17.2; DB 6; Length 470;
XX Best Local Similarity 86.4%; Pred. No. 1.7e+02;
XX Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GATCTTGACGTGCTTAACCC 22
DB 292 GATCTTGTTGTCATTAACCC 313

RESULT 5
AAC5036/C
ID AAC5036 standard; DNA; 636 BP.
XX
XX AAC5036;
XX
XX 18-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 79959.
XX
XX Hybridisation assay; genetic mapping; gene expression control;
XX protein identification; signal transduction pathway; metabolic pathway;
XX promoter; termination sequence; ss.
XX
XX Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-00301439.
XX
XX 25-FEB-1999; 99US-0121825P.
XX 05-MAR-1999; 99US-0123180P.
XX 09-MAR-1999; 99US-0123548P.
XX 23-MAR-1999; 99US-0125788P.
XX 25-MAR-1999; 99US-0126264P.
XX 29-MAR-1999; 99US-0126785P.
XX 01-APR-1999; 99US-0127462P.
XX 06-APR-1999; 99US-0128234P.
XX 08-APR-1999; 99US-0128714P.
XX 16-APR-1999; 99US-0129845P.
XX 19-APR-1999; 99US-0130077P.
XX 21-APR-1999; 99US-0130449P.
XX 23-APR-1999; 99US-0130510P.
XX 23-APR-1999; 99US-0130891P.
XX 28-APR-1999; 99US-0131449P.
XX 30-APR-1999; 99US-0132048P.
XX 30-APR-1999; 99US-0132407P.
XX 04-MAY-1999; 99US-0132484P.
XX 05-MAY-1999; 99US-0132485P.
XX 06-MAY-1999; 99US-0132486P.
XX 06-MAY-1999; 99US-0132487P.
```



Wed Feb 9 09:28:16 2005

PR 07-MAY-1999; 99US-0132863P.  
PR 11-MAY-1999; 99US-0134256P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135529P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 28-JUN-1999; 99US-0140921P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 15-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 26-JUL-1999; 99US-0145224P.  
PR 27-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 22-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151103P.  
PR 31-AUG-1999; 99US-0151138P.  
PR 01-SEP-1999; 99US-015130P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 16-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.

PR 22-OCT-1999; 99US-0160980P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160982P.  
PR 25-OCT-1999; 99US-0161404P.  
PR 25-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 25-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 26-OCT-1999; 99US-0161361P.  
PR 28-OCT-1999; 99US-0161920P.  
PR 28-OCT-1999; 99US-0161932P.  
PR 28-OCT-1999; 99US-0161933P.  
PR 29-OCT-1999; 99US-0162142P.

Query Match 78.2%; Score 17.2; DB 3; Length 636;  
Best Local Similarity 86.4%; Pred. No. 1.8e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 GATCTGTACTGCTTAACCC 22  
Db 363 GATCTGTCTTGTGATTAACCC 342

RESULT 6  
AAC53341/c  
ID AAC53341 standard; DNA; 641 BP.

AC AAC53341;  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana DNA fragment SRQ ID NO: 74193.  
XX  
KW Hybridisation assay; genetic mapping; gene expression control;  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW Promoter; termination sequence; ss.

XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-00301439.  
XX  
PR 25-FEB-1999; 99US-0121825P.  
PR 05-MAR-1999; 99US-0123180P.  
PR 09-MAR-1999; 99US-0123548P.  
PR 23-MAR-1999; 99US-0125788P.  
PR 25-MAR-1999; 99US-0126264P.  
PR 29-MAR-1999; 99US-0126785P.  
PR 01-APR-1999; 99US-0127462P.  
PR 06-APR-1999; 99US-0128234P.  
PR 08-APR-1999; 99US-0128714P.  
PR 16-APR-1999; 99US-0129845P.  
PR 19-APR-1999; 99US-0130077P.  
PR 21-APR-1999; 99US-0130449P.  
PR 23-APR-1999; 99US-0130510P.  
PR 23-APR-1999; 99US-0130891P.  
PR 28-APR-1999; 99US-0131449P.  
PR 30-APR-1999; 99US-0132048P.  
PR 30-APR-1999; 99US-0132407P.  
PR 04-MAY-1999; 99US-0132484P.  
PR 05-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 06-MAY-1999; 99US-0132487P.  
PR 07-MAY-1999; 99US-0132633P.  
PR 11-MAY-1999; 99US-0134256P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134321P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.

PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144332P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.

PR 28-JUL-1999; 99US-0145951P.  
 PR 02-AUG-1999; 99US-0146386P.  
 PR 02-AUG-1999; 99US-0146388P.  
 PR 02-AUG-1999; 99US-0146389P.  
 PR 03-AUG-1999; 99US-0147038P.  
 PR 04-AUG-1999; 99US-0147204P.  
 PR 04-AUG-1999; 99US-0147302P.  
 PR 05-AUG-1999; 99US-0147392P.  
 PR 05-AUG-1999; 99US-0147260P.  
 PR 06-AUG-1999; 99US-0147303P.  
 PR 06-AUG-1999; 99US-0147416P.  
 PR 09-AUG-1999; 99US-0147493P.  
 PR 09-AUG-1999; 99US-0147935P.  
 PR 10-AUG-1999; 99US-0148171P.  
 PR 11-AUG-1999; 99US-0148319P.  
 PR 12-AUG-1999; 99US-0148341P.  
 PR 13-AUG-1999; 99US-0148565P.  
 PR 13-AUG-1999; 99US-0148684P.  
 PR 16-AUG-1999; 99US-0149368P.  
 PR 17-AUG-1999; 99US-0149175P.  
 PR 18-AUG-1999; 99US-0149426P.  
 PR 20-AUG-1999; 99US-0149722P.  
 PR 20-AUG-1999; 99US-0149723P.  
 PR 20-AUG-1999; 99US-0149929P.  
 PR 23-AUG-1999; 99US-0149902P.  
 PR 23-AUG-1999; 99US-0149930P.  
 PR 25-AUG-1999; 99US-0150566P.  
 PR 26-AUG-1999; 99US-0150884P.  
 PR 27-AUG-1999; 99US-0151065P.  
 PR 27-AUG-1999; 99US-0151066P.  
 PR 27-AUG-1999; 99US-0151080P.  
 PR 30-AUG-1999; 99US-0151303P.  
 PR 31-AUG-1999; 99US-0151438P.  
 PR 01-SEP-1999; 99US-0151930P.  
 PR 07-SEP-1999; 99US-0152363P.  
 PR 10-SEP-1999; 99US-0153070P.  
 PR 13-SEP-1999; 99US-0153758P.  
 PR 15-SEP-1999; 99US-0154018P.  
 PR 16-SEP-1999; 99US-0154039P.  
 PR 20-SEP-1999; 99US-0154779P.  
 PR 22-SEP-1999; 99US-0155139P.  
 PR 23-SEP-1999; 99US-0155486P.  
 PR 24-SEP-1999; 99US-0155659P.  
 PR 28-SEP-1999; 99US-0156458P.  
 PR 29-SEP-1999; 99US-0156586P.  
 PR 04-OCT-1999; 99US-0157117P.  
 PR 05-OCT-1999; 99US-0157753P.  
 PR 06-OCT-1999; 99US-0157665P.  
 PR 07-OCT-1999; 99US-0158029P.  
 PR 08-OCT-1999; 99US-0158232P.  
 PR 12-OCT-1999; 99US-0158369P.  
 PR 13-OCT-1999; 99US-0159293P.  
 PR 13-OCT-1999; 99US-0159294P.  
 PR 13-OCT-1999; 99US-0159295P.  
 PR 14-OCT-1999; 99US-0159339P.  
 PR 14-OCT-1999; 99US-0159330P.  
 PR 14-OCT-1999; 99US-0159331P.  
 PR 14-OCT-1999; 99US-0159637P.  
 PR 14-OCT-1999; 99US-0159638P.  
 PR 18-OCT-1999; 99US-0159584P.  
 PR 21-OCT-1999; 99US-0160741P.  
 PR 21-OCT-1999; 99US-0160767P.  
 PR 21-OCT-1999; 99US-0160768P.  
 PR 21-OCT-1999; 99US-0160770P.  
 PR 21-OCT-1999; 99US-0160814P.  
 PR 21-OCT-1999; 99US-0160815P.  
 PR 22-OCT-1999; 99US-0160980P.  
 PR 22-OCT-1999; 99US-0160981P.  
 PR 22-OCT-1999; 99US-0160989P.  
 PR 25-OCT-1999; 99US-0161404P.  
 PR 25-OCT-1999; 99US-0161405P.  
 PR 25-OCT-1999; 99US-0161406P.  
 PR 26-OCT-1999; 99US-0161359P.

PR 26-OCT-1999; 99US-0161360P.  
 PR 26-OCT-1999; 99US-0161361P.  
 PR 28-OCT-1999; 99US-0161320P.  
 PR 28-OCT-1999; 99US-0161922P.  
 PR 28-OCT-1999; 99US-0161933P.  
 PR 29-OCT-1999; 99US-0162142P.

Query Match 78.2%; Score 17.2; DB 3; Length 641;  
 Best Local Similarity 86.4%; Pred. No. 1.8e+02;  
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Db 365 GATCTGTGTTTGTCTAAACC 344  
 1 GATCTGTGTTTGTCTAAACC 22  
 ||||| ||||| ||||| |||||

RESULT 7  
 ID ADP13274/c  
 ADP13274 standard; DNA; 101569 BP.

AC ADP13274;  
 DT 26-AUG-2004 (first entry)  
 XX  
 DE Renal cell carcinoma differentially expressed gene #10.  
 XX  
 KW de; diagnosis; non-blood disease; solid tumor; gene expression;  
 KW peripheral blood mononuclear cell; renal cell carcinoma; prostate cancer;  
 KW head/neck cancer; differential expression.  
 XX  
 OS Homo sapiens.  
 XX  
 .FN WO2004048933-A2.  
 PD 10-JUN-2004.  
 XX  
 PP 21-NOV-2003; 2003MO-US037481.  
 PR 21-NOV-2002; 2002US-0427982P.  
 PR 03-APR-2003; 2003US-0459782P.

XX (AMHP ) WYETH.  
 PA (TWIN/) TWINE N C.  
 PA (BURC/) BURCZYNSKI M E.  
 PA (TREP/) TREPICCHIO W L.  
 PA (DORN/) DORNER A.  
 PA (STOV/) STOVER J A.  
 PA (SLON/) SLONI D K.

PI Twine NC, Burczynski ME, Trepicchio WL, Dorner A, Stover JA;  
 PI Sloni DK;  
 XX  
 DR WPI; 2004-460799/43.

XX  
 PT diagnosing non-blood disease such as solid tumor, involves comparing  
 PT differential expression profile of specific genes in peripheral blood  
 PT sample of subject with reference expression profile of specific genes.  
 XX  
 XX  
 XX Disclosure; SEQ ID NO 10; 350bp; English.

CC The invention relate to a method of diagnosing (M1) non-blood disease  
 CC such as solid tumor by providing peripheral blood sample of human having  
 CC non-blood disease, and comparing an expression profile of specific genes  
 CC in the peripheral blood sample to reference expression profile of the  
 CC genes, where each of the genes is differentially expressed in peripheral  
 CC blood mononuclear cells (PBMCs) of patients having the disease as  
 CC compared to PBMCs of normal humans. The method is useful for diagnosing  
 CC non-blood disease such as solid tumor. The solid tumor is chosen from  
 CC renal cell carcinoma (RCC), prostate cancer and head/neck cancer. The  
 CC peripheral blood sample comprises enriched PBMCs. The peripheral blood  
 CC sample is a whole blood sample (claimed). (M1) is useful for identifying  
 CC genes that are differentially expressed in peripheral blood samples  
 CC isolated at different stages of progression, development or treatment of

CC RCC and/or other solid tumors. This sequence corresponds to a gene that  
CC is differentially expressed and detected by the method of the invention.  
CC (Note: this sequence is not given as part of the printed specification  
CC but was obtained from WIPO in electronic format at  
CC ftp.wipo./pub/published\_pct\_sequences).

XX  
SQ Sequence 101569 BP; 27935 A; 22428 C; 22623 G; 28582 T; 0 U; 1 Other;

Query Match 78.2%; Score 17.2; DB 12; Length 101569;  
Best Local Similarity 86.4%; Pred. No. 3e+02;

Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 GATCTTGACTGCTTTAAACCC 22  
Db 55369 GATCTTGACTGCTTTAAACCC 55348

RESULT 8  
ADQ18492/c  
ID ADQ18492 standard; DNA; 259202 BP.

XX  
AC ADQ18492;

XX  
DT 26-AUG-2004 (first entry)

XX  
DE Human soft tissue sarcoma-upregulated DNA - SEQ ID 1311.

XX  
KM soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;  
KM dr.

XX  
OS Homo sapiens.

XX  
PN WO2004048938-A2.

XX  
PD 10-JUN-2004.

XX  
PF 26-NOV-2003; 2003WO-US038193.

XX  
PR 26-NOV-2002; 2002US-0429739P.

XX  
PA (PROT-) PROTEIN DESIGN LABS INC.

XX  
PI Aziz N, Ginsburg WM, Zlotnick A;

XX  
DR WPI; 2004-441208/41.

PT Early detection of soft tissue sarcoma comprises determining expression  
PT of a gene in a first soft tissue sample and a normal soft tissue sample  
PT and comparing the gene expression, also useful in treating soft tissue  
PT sarcoma.

XX  
PS Example 2; SEQ ID NO 1311; 210pp; English.

CC The invention relates to a novel method for detecting soft tissue sarcoma  
CC which comprises obtaining a first soft tissue sample from an individual  
CC and a normal soft tissue sample from the same or different individual,  
CC determining the expression of a gene in both samples and comparing the  
CC expression of the gene in both soft tissue samples, where a higher level  
CC of protein expression in the first soft tissue sample indicates the  
CC presence of soft tissue sarcoma. The method of the invention has  
CC cytostatic applications and may be useful for detecting soft tissue  
CC sarcoma, possibly via gene therapy or vaccine production. The nucleic  
CC acid sequences may be useful in diagnostic and screening applications.  
CC The current sequence is that of a human soft tissue sarcoma-upregulated  
CC DNA of the invention. The current sequence is not shown within the  
CC specification per se but was submitted in CD format by the inventor.

XX  
SQ Sequence 259202 BP; 81699 A; 51421 C; 49221 G; 76861 T; 0 U; 0 Other;

Query Match 78.2%; Score 17.2; DB 12; Length 259202;  
Best Local Similarity 86.4%; Pred. No. 3.2e+02;

Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 GATCTTGACTGCTTTAAACCC 22

Db 70030 GATCTTGACTGCTTTAAACCC 70009

RESULT 9

ABN78623/c  
ID ABN78623 standard; cDNA; 567 BP.

XX  
AC ABN78623;

XX  
DT 08-JUL-2002 (first entry)

XX  
DE Human synthase-like ORF3570 cDNA, SEQ ID NO:7139.

XX Human; ORF, open reading frame; ORFX; drug screening; diagnosis;  
XX disease monitoring; cytokine; cell proliferation; cell differentiation;  
XX immune modulation; haematopoiesis regulation; tissue growth;  
XX angiogenesis; activin; inhibin; chemotactic; chemokinetic; haemostatic;  
XX thrombolytic; tumour inhibition; bodily characteristics; fertility;  
XX behaviour; cancer; proliferative disorder; neurological disorder;  
XX cardiovascular disease; immune system disorder; organ transplantation;  
XX tissue growth disorder; tissue regeneration disorder; diabetes mellitus;  
XX hypothyroidism; cholesterol ester storage disease; infection; vulnery;  
XX vasotropic; antiposrotic; antidiabetic; cyostatic; nootropic;  
XX neuroprotective; antiatherosclerotic; anticoagulant; thrombolytic;  
XX cardiant; hypotensive; antihyroid; antiinflammatory; immunomodulator;  
XX dermatological; analgesic; virucide; antibacterial; fungicide; gene; ss.

XX  
OS Homo sapiens.

XX  
PN WO200190366-A2.

XX  
PD 29-NOV-2001.

XX  
PF 24-MAY-2001; 2001WO-US017076.

XX  
PR 24-MAY-2000; 2000US-0206690P.

XX  
PA (CUPRA-) CUPAGEN CORP.

XX  
PI Leach MD, Shinkels RA;

XX  
DR WPI; 2002-106200/14.

XX  
PT P-PSDB; ABP34597.

PT Novel human polypeptides and polynucleotides useful for diagnosing,  
PT preventing and treating cardiovascular disease, neurodegenerative,  
PT hyperproliferative disorders and disorders related to organ  
PT transplantation.

XX  
PS Claim 1; Page 2031; 2508pp; English.

XX Sequences ABP1028-ABP35561 represent 4534 novel human proteins  
XX designated ORF (open reading frame) 1-4534, and sequences ABN75054-  
XX ABN79587 represent cDNAs encoding them. The invention also encompasses  
XX polypeptides at least 80% identical to the ORF1-ORF4534 (collectively  
XX referred to as ORFX) proteins, polynucleotides at least 85% identical to  
XX the ORFX nucleic acid sequences, vectors and host cells comprising ORFX  
XX polynucleotides, the recombinant production of ORFX proteins, antibodies  
XX specific for ORFX proteins, methods of detecting ORFX polynucleotides and  
XX polypeptides, methods of screening for modulators of ORFX expression or  
XX activity, and methods of screening individuals for a predisposition to an  
XX ORFX-associated disorder. The ORFX proteins of the invention have a wide  
XX range of biological activities, such as cytokine, cell proliferation,  
XX cell differentiation, immune modulation, haematopoiesis regulation,  
XX tissue growth, angiogenesis, activin or inhibin activity, chemotactic/  
XX chemokinetic activity, haemostatic activity, thrombolytic activity,  
XX receptor/ligand, antiinflammatory activity, tumour inhibition activity,  
XX and antifibrotic activity, and may also be involved in the determination  
XX of bodily characteristics, fertility and behaviour. ORFX proteins,  
XX nucleic acids and antibodies may be used in the treatment of cancers,  
XX other proliferative disorders such as psoriasis and benign tumours,

CC neurological disorders such as epilepsy and Alzheimer's disease.  
CC cardiovascular diseases, immune system disorders, disorders related to  
CC organ transplantation, disorders of tissue growth and regeneration,  
CC diseases such as diabetes mellitus, hypothyroidism, and cholesterol  
CC storage disease, and infectious diseases caused by viral, bacterial,  
CC fungal and other pathogens. ORFX nucleic acids may also be used as a  
CC source of primers and probes, in the detection of ORFX genomic sequences  
CC or transcripts, in the identification and cloning of homologous  
CC sequences, in genetic diagnosis, and in forensic biology. The ORFX  
CC nucleic acids may additionally be used to produce transgenic animals  
CC which may be useful for studying the function and/or activity of ORFX  
CC protein, and in drug screening. The ORFX proteins may also be used as  
CC immunogens to generate specific antibodies, which are useful in the  
CC diagnosis, treatment and monitoring of ORFX-associated diseases

XX Sequence 567 BP; 146 A; 97 C; 133 G; 191 T; 0 U; 0 Other;

Query Match 76.4%; Score 16.8; DB 6; Length 567;  
Best Local Similarity 90.0%; Pred. No. 2.8e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 TCTTGTACTGCTTTAAACC 22  
DB 303 TCTGTATTGTCTGTAACCC 284

RESULT 10  
ABN61229/c  
ID ABN61229 standard; cDNA; 637 BP.  
XX  
AC ABN61229;  
XX  
DT 28-JUN-2002 (first entry)  
XX  
DE Human cancer related polynucleotide SEQ ID NO 1196.  
XX  
KM Human; cytostatic; gene expression; gene mapping; tissue profiling;  
XX Gene therapy; cancer; tumor; gene; 88.  
XX  
OS Homo sapiens.  
XX  
PN WO200214500-A2.  
XX  
PD 21-FEB-2002.  
XX  
PF 16-AUG-2001; 2001WO-US025840.  
XX  
PR 16-AUG-2000; 2000US-0226326P.  
XX  
PA (CHIR) CHIRON CORP.  
XX (HYSE-) HYSEQ INC.  
XX  
PI Sacobedo J, Garcia PD, Sudduth-Klinger J, Reinhard C, Randazzo F;  
XX Lameon G, Scott EM, Zhang G, Kassam A, Pot D, Labat I;  
XX WPI; 2002-241905/29.  
XX  
PT New nucleic acid for producing a polypeptide, detecting differentially  
XX expressed genes correlated with a cancerous state of a mammalian cell,  
XX and inhibiting tumor growth.  
XX  
PS Claim 1; SEQ ID NO 1196; 883pp + Sequence Listing; English.

CC The invention relates to an isolated polynucleotide (ABN27253-ABN33262)  
CC with cytostatic activity. The polynucleotide is used to produce a  
CC polypeptide, to detect differentially expressed genes correlated with a  
CC cancerous state of a mammalian cell and to inhibit tumor growth. The  
CC polynucleotide is used as a probe in mapping and tissue profiling. The  
CC encoded polypeptide and antibodies to the polypeptide can also be used  
CC for therapeutic and diagnostic purposes. The polynucleotide is useful for  
CC gene therapy. Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic format  
CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 637 BP; 265 A; 84 C; 122 G; 166 T; 0 U; 0 Other;

QY 3 TCTTGTACTGCTTTAAACC 22  
DB 543 TCTGTACTGCTTTCAACC 524

RESULT 11  
ACC60893/c  
ID ACC60893 standard; DNA; 1452 BP.  
XX  
AC ACC60893;  
XX  
DT 20-JUN-2003 (first entry)  
XX  
DE Gene sequence #SEQ ID 568.  
XX  
KM Multiprotein complex; eukaryote; drug target; diagnosis; gene; ds.  
XX  
OS Saccharomyces cerevisiae.  
XX  
PN EP1258494-A1.  
XX  
PD 20-NOV-2002.  
XX  
PF 20-DEC-2001; 2001EP-00130253.  
XX  
PR 15-MAY-2001; 2001EP-0011774.  
XX  
PA (CELL-) CELLZONE AG.  
XX  
PI Bauer A, Gavin A, Grandi P, Krause R, Kruse UD, Kuester BD;  
XX Marzloch M, Schultz JD, Superti-Furga GD;  
XX WPI; 2003-250078/25.  
XX  
DR P-PDB; ABR52851.  
XX  
PT New isolated protein complexes useful for diagnosing a disease or  
XX disorder, or as a target for an active agent of a pharmaceutical,  
XX preferably a drug target in the treatment or prevention of disease or  
XX disorder.  
XX  
PS Disclosure; SEQ ID NO 568; 17pp + Sequence Listing; English.

CC The invention relates to multiprotein complexes from eukaryotes. Proteins  
CC of the invention and DNA sequences encoding them are given in records  
CC ABR52568-ABR53903 and ACC60610-ACC61944 respectively. The complexes are  
CC obtainable by using a protein as a bait and isolating the set of proteins  
CC which is attached thereto from cells. Such protein complexes may comprise  
CC up to 30 distinct proteins. Protein complexes of the invention are useful  
CC for diagnosing a disease or disorder, or as a target for an active agent  
CC of a pharmaceutical, preferably a drug target in the treatment or  
CC prevention of a disease or disorder. Note: The sequence data for this  
CC patent is not represented in the printed specification, but is based on  
CC sequence information supplied by the European Patent Office. The complete  
CC document is available on CD-ROM

XX Sequence 1452 BP; 489 A; 226 C; 351 G; 386 T; 0 U; 0 Other;

Query Match 76.4%; Score 16.8; DB 10; Length 1452;  
Best Local Similarity 90.0%; Pred. No. 3e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 TCTTGTACTGCTTTAAACC 22  
DB 741 TCTGTATTGTCTGTAACCC 722

## RESULT 12

ADK62303/C

ID ADK62303 standard; DNA; 1452 BP.

AC ADK62303;

DT 06-MAY-2004 (first entry)

DE Disease treating protein complex-derived gene #277.

XX protein complex; drug target; diagnosis; gene; ds.

XX Unidentified.

XX EPI338608-A2.

XX 27-AUG-2003.

XX 20-DEC-2002; 2002EP-00102902.

XX 20-DEC-2001; 2001EP-00130253.

XX (CELL-) CELLZONE AG.

XX Bauer A, Gavin A, Superti-Furga G, Kuester B, Schultz J;

XX Merzloch M, Grandt P, Krause R, Kruse U, Merino A, Bauch A;

XX Michon A, Leutwein C, Rick J;

XX WPI; 2003-638460/61.

XX P-PSDB; ADK62302.

XX New proteins and protein complexes from eukaryotes, useful as targets in

XX drug screening, or in diagnosing or screening for the presence of a

XX disease or disorder, or a predisposition for developing a disease or

XX disorder in a subject.

XX Disclosure; SEQ ID NO 554; 13bp; English.

XX The invention relates to novel protein complexes comprising a first and a

XX second protein, or its derivative, fragment, homologue or variant. The

XX proteins are selected from given protein complexes, which are not defined

XX in the specification. The variants are encoded by nucleic acids that

XX hybridize to the nucleic acids encoding the proteins under low stringency

XX conditions. The protein complexes are useful as targets for an active

XX agent of a pharmaceutical. These protein complexes are particularly

XX useful as drugs targets for the treatment or preventing of a disease or

XX disorder. The complexes and methods above are useful in diagnosing or

XX screening for the presence of a disease or disorder or a predisposition

XX for developing a disease or disorder in a subject. These are also useful

XX in screening for a drug for treatment or prevention of a disease or

XX disorder. The molecule that modulates the amount, activity or protein

XX components of the complex is useful for the manufacture of a medicament

XX for the treatment or prevention of a disease or disorder. This sequence

XX corresponds to a gene of the invention. (Note: the sequence data for this

XX patent did not form part of the printed specification but was obtained

XX from the EPO in electronic format).

XX Sequence 1452 BP; 489 A; 226 C; 351 G; 386 T; 0 U; 0 Other;

XX Query Match 76.4%; Score 16.8; DB 10; Length 1452;

XX Best Local Similarity 90.0%; Pred. No. 3e+02;

XX Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

XX 3 TCTTGTACTGCTTAAACC 22

XX TCTTGTACTGCTTAAACC 722

XX TCTTGTACTGCTTAAACC 722

XX TCTTGTACTGCTTAAACC 722

XX TCTTGTACTGCTTAAACC 722

XX TCTTGTACTGCTTAAACC 722

XX 06-OCT-2000 (first entry)

XX Human secreted protein 5' EST, SEQ ID NO: 33503.

XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;

XX gene therapy; chromosome mapping; ss.

XX Homo sapiens.

XX EPI033401-A2.

XX 06-SEP-2000.

XX 21-FEB-2000; 2000EP-00200610.

XX 26-FEB-1999; 99US-0122487P.

XX (GEST) GENSET.

XX Dumae Malne Edwards J, Duclert A, Giordano J;

XX WPI; 2000-500381/45.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for

XX obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for

XX diagnostic, forensic, gene therapy and chromosome mapping procedures.

XX Claim 1, SEQ ID NO 33503; 71bp + Sequence listing; English.

XX The present sequence is one of a large number of 5' ESTs derived from

XX cDNAs encoding secreted proteins. No ORF has yet been conclusively

XX identified within the present sequence. The 5' ESTs were prepared from

XX total human RNAs or polyA+ RNAs derived from 30 different tissues. EST

XX sequences usually correspond mainly to the 3' untranslated region (UTR)

XX of the mRNA because they are often obtained from oligo-dT primed cDNA

XX libraries. Such ESTs are not well suited for isolating cDNA sequences

XX derived from the 5' ends of mRNAs and even in those cases where longer

XX cDNA sequences have been obtained, the full 5' UTR is rarely included. 5'

XX ESTs are derived from mRNAs with intact 5' ends and can therefore be used

XX to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in

XX diagnostic, forensic, gene therapy and chromosome mapping procedures.

XX They are used to obtain upstream regulatory sequences and to design

XX expression and secretion vectors

XX Sequence 356 BP; 93 A; 81 C; 52 G; 130 T; 0 U; 0 Other;

XX Query Match 74.5%; Score 16.4; DB 3; Length 356;

XX Best Local Similarity 94.4%; Pred. No. 4.1e+02;

XX Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

XX 3 TCTTGTACTGCTTAAAC 20

XX TCTTGTACTGCTTAAAC 349

XX TCTTGTACTGCTTAAAC 349

XX TCTTGTACTGCTTAAAC 349

XX TCTTGTACTGCTTAAAC 349

## RESULT 14

ABA60565

ID ABA60565 standard; DNA; 508 BP.

AC ABA60565;

DT 01-FEB-2002 (first entry)

DE Human foetal liver single exon nucleic acid probe #8870.

XX Human; foetal liver; gene expression; single exon nucleic acid probe; ss.

XX Homo sapiens.

XX WO200157277-A2.

XX 09-AUG-2001.

```

XX PF 30-JAN-2001; 2001WO-US000669.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI, 2001-483447/52.
XX PT Human genome-derived single exon nucleic acid probes useful for analyzing
XX PT gene expression in human fetal liver.
XX PS Claim 1; SEQ ID NO 8870; 639pp + Sequence Listing; English.
XX CC The invention relates to a single exon nucleic acid probe for measuring
XX CC human gene expression in a sample derived from human foetal liver. The
XX CC single exon nucleic acid probes may be used for predicting, measuring and
XX CC displaying gene expression in samples derived from human fetal liver. The
XX CC present sequence is a single exon nucleic acid probe of the invention.
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 508 BP; 131 A; 118 C; 95 G; 164 T; 0 U; 0 Other;

QY Query Match 74.5%; Score 16.4; DB 4; Length 508;
Best Local Similarity 94.4%; Pred. No. 4.3e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 3 TCTTGACTGTCTTAAC 20
255 TCTTTACTGTCTTAAC 272

RESULT 15
AA140454
ID AA140454 standard; DNA; 508 BP.
XX AC AA140454;
XX DT 17-OCT-2001 (first entry)
XX DE Probe #9140 used to measure gene expression in human placenta sample.
XX KW Probe; microarray; human; placenta; antenatal diagnosis;
XX KW genetic disorder; ss.
XX OS Homo sapiens.
XX PN WO200157272-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US000663.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX

```

```

PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI, 2001-48897/53.
XX PT Human genome-derived single exon nucleic acid probes useful for analyzing
XX PT gene expression in human placenta.
XX PS Claim 25; SEQ ID NO 9140; 654pp; English.
XX CC The present invention relates to single exon nucleic acid probes (SNP).
XX CC The present sequence is one such probe. The probes are useful for
XX CC producing a microarray for predicting, measuring and displaying gene
XX CC expression in samples derived from human placenta. The probes are useful
XX CC for antenatal diagnosis of human genetic disorders
XX SQ Sequence 508 BP; 131 A; 118 C; 95 G; 164 T; 0 U; 0 Other;

QY Query Match 74.5%; Score 16.4; DB 4; Length 508;
Best Local Similarity 94.4%; Pred. No. 4.3e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 3 TCTTGACTGTCTTAAC 20
255 TCTTTACTGTCTTAAC 272

Search completed: February 8, 2005, 13:56:56
Job time : 34.1839 secs

```

**This Page Blank (uspto)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 8, 2005, 13:03:55 / Search time 5.94179 Seconds

(Without alignments)  
2631.757 Million cell updates/sec

Title: US-10-754-437-5

Perfect score: 22  
Sequence: 1 gactctgtactgtcttaacc 22

Scoring table: IDENTITY\_NUC

Gapop 10.0, Gapext 1.0

Searched: 824507 seqs, 355394441 residues

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database:

Issued Patents NA: \*  
1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq: \*  
2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq: \*  
3: /cgn2\_6/ptodata/1/ina/6A.COMB.seq: \*  
4: /cgn2\_6/ptodata/1/ina/6B.COMB.seq: \*  
5: /cgn2\_6/ptodata/1/ina/PTUS.COMB.seq: \*  
6: /cgn2\_6/ptodata/1/ina/backfile1.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	17.2	78.2	479	4 US-09-621-976-3622	Sequence 3622, Ap
2	16.4	74.5	356	4 US-09-513-999C-33503	Sequence 33503, A
3	16.2	73.6	199	4 US-09-313-394A-3083	Sequence 3083, Ap
4	16.2	73.6	279	4 US-09-248-796A-8060	Sequence 8060, Ap
5	16.2	73.6	388	3 US-08-998-416-960	Sequence 960, App
6	16.2	73.6	5554	4 US-09-815-923-1	Sequence 1, App1
7	16	72.7	741	4 US-09-543-681A-4040	Sequence 4040, App
8	15.8	71.8	567	4 US-09-107-532A-618	Sequence 618, App
9	15.6	70.9	346	4 US-09-513-999C-10254	Sequence 10254, A
10	15.6	70.9	1039	4 US-09-566-921-2	Sequence 2, App1
11	15.6	70.9	1980	4 US-09-543-681A-576	Sequence 576, App
12	15.4	70.0	1251	4 US-10-164-595-25	Sequence 25, App1
13	15.4	70.0	1341	4 US-08-956-171E-448	Sequence 448, App
14	15.4	70.0	1341	4 US-08-781-986A-448	Sequence 448, App
15	15.4	70.0	14770	4 US-09-230-132-30	Sequence 30, App
16	15.4	70.0	392000	4 US-10-027-983-11	Sequence 11, App1
17	15.2	69.1	306	2 US-08-634-797-16	Sequence 16, App1
18	15.2	69.1	306	2 US-08-634-797-28	Sequence 28, App1
19	15.2	69.1	331	4 US-09-513-999C-26343	Sequence 26343, A
20	15.2	69.1	524	4 US-09-270-767-11178	Sequence 11178, A
21	15.2	69.1	528	4 US-09-583-110-1687	Sequence 1687, Ap
22	15.2	69.1	660	4 US-09-248-796A-4830	Sequence 4830, Ap
23	15.2	69.1	1020	4 US-09-107-532A-1860	Sequence 1860, Ap
24	15.2	69.1	1146	4 US-09-489-039A-985	Sequence 985, App1
25	15.2	69.1	1296	4 US-08-714-741-43	Sequence 43, App1
26	15.2	69.1	1587	4 US-09-248-796A-3998	Sequence 3998, Ap
27	15.2	69.1	2059	4 US-08-714-741-45	Sequence 45, App1

C 28	15.2	69.1	3023	3 US-09-308-022-4	Sequence 4, App1
C 29	15.2	69.1	3222	4 US-08-714-741-39	Sequence 39, App1
C 30	15.2	69.1	1664976	4 US-08-916-421B-1	Sequence 1, App1
C 31	15.2	69.1	1664976	4 US-09-692-570-1	Sequence 1, App1
C 32	15	68.2	10419	4 US-09-408-020-3	Sequence 3, App1
C 33	15	68.2	42432	4 US-09-408-020-2	Sequence 2, App1
C 34	14.8	67.3	263	4 US-09-513-999C-20021	Sequence 20021, A
C 35	14.8	67.3	375	4 US-09-248-796A-12298	Sequence 12298, A
C 36	14.8	67.3	423	1 US-08-470-179-150	Sequence 150, App
C 37	14.8	67.3	1189	4 US-09-799-451-709	Sequence 709, App
C 38	14.8	67.3	1833	4 US-09-248-796A-2927	Sequence 2927, App
C 39	14.8	67.3	2084	4 US-09-222-938A-9	Sequence 9, App1
C 40	14.8	67.3	8054	4 US-09-581-909-2	Sequence 2, App1
C 41	14.8	67.3	28626	4 US-09-586-002-15	Sequence 15, App1
C 42	14.8	67.3	1830121	4 US-09-557-884-1	Sequence 1, App1
C 43	14.8	67.3	1830121	4 US-09-643-990A-1	Sequence 1, App1
C 44	14.8	67.3	1830121	4 US-10-329-960-1	Sequence 1, App1
C 45	14.6	66.4	347	4 US-09-621-976-12653	Sequence 12653, A

# ALIGNMENTS

```

RESULT 1
US-09-621-976-3622
Sequence 3622, Application US/09621976
Patent No. 6639063
GENERAL INFORMATION:
APPLICANT: Dumas Milne Edwards, J.B.
APPLICANT: Jobert, S.
TITLE OF INVENTION: ESTs and Encoded Human Proteins.
FILE REFERENCE: GENSET.054PR2
CURRENT APPLICATION NUMBER: US/09/621,976
CURRENT FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 19335
SOFTWARE: Patent.pm
SEQ ID NO 3622
LENGTH: 479
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: 133..282
US-09-621-976-3622

Query Match          78.2% Score 17.2; DB 4; Length 479;
Best local similarity 86.4%; Pred. No. 18;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 GATCTGTACTGTCTTAACCC 22
Db      421 GATCTGTACTGTCTTAACCC 442

RESULT 2
US-09-513-999C-33503
Sequence 33503, Application US/09513999C
Patent No. 6783961
GENERAL INFORMATION:
APPLICANT: Dumas Milne Edwards, J.B.
APPLICANT: Duclerc, A.
APPLICANT: Giordano, J.Y.
TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
FILE REFERENCE: 59 US2, REG
CURRENT APPLICATION NUMBER: US/09/513,999C
CURRENT FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/122,487
PRIOR FILING DATE: 1999-02-26
NUMBER OF SEQ ID NOS: 36681
SOFTWARE: Patent.pm
SEQ ID NO 33503

```

LENGTH: 356  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-513-999C-33503

Query Match 74.5%; Score 16.4; DB 4; Length 356;  
Best Local Similarity 94.4%; Pred. No. 44;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 TCTTGTACTGCTTAAAC 20  
DB 332 TCTTGTACTGCTTAAAC 349

RESULT 3  
US-09-313-294A-3083  
Sequence 3083, Application US/09313294A  
Patent No. 6476212  
GENERAL INFORMATION:  
APPLICANT: Laljudi, Raghunath V.  
APPLICANT: Ito, Laura Y.  
APPLICANT: Sherman, Bradley K.  
TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR  
FILE REFERENCE: PL-0017 US  
CURRENT APPLICATION NUMBER: US/09/313,294A  
CURRENT FILING DATE: 1999-05-14  
NUMBER OF SEQ ID NOS: 7600  
SOFTWARE: PERL Program  
SEQ ID NO 3083  
LENGTH: 199  
TYPE: DNA  
ORGANISM: Zea mays  
FEATURE:  
NAME/KEY: misc\_feature  
OTHER INFORMATION: Incyte ID No. 6476212 700610840H1  
NAME/KEY: unsure  
LOCATION: 6, 86  
OTHER INFORMATION: a, t, c, g, or other  
US-09-313-294A-3083

Query Match 73.6%; Score 16.2; DB 4; Length 199;  
Best Local Similarity 85.7%; Pred. No. 51;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 ATCTGTACTGCTTAAACC 22  
DB 8 ATCTGTACTGCTTAAACC 28

RESULT 4  
US-09-248-796A-8060  
Sequence 8060, Application US/09248796A  
Patent No. 6747137  
GENERAL INFORMATION:  
APPLICANT: Keith Weinstock et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN  
TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 107196.132  
CURRENT APPLICATION NUMBER: US/09/248,796A  
CURRENT FILING DATE: 1999-02-12  
PRIOR APPLICATION NUMBER: US 60/074,725  
PRIOR FILING DATE: 1998-02-13  
PRIOR APPLICATION NUMBER: US 60/096,409  
PRIOR FILING DATE: 1998-08-13  
NUMBER OF SEQ ID NOS: 28208  
SEQ ID NO 8060  
LENGTH: 279  
TYPE: DNA  
ORGANISM: Candida albicans  
US-09-248-796A-8060

Query Match 73.6%; Score 16.2; DB 4; Length 279;  
Best Local Similarity 85.7%; Pred. No. 53;

Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 ATCTGTACTGCTTAAACC 22  
DB 43 ATCTGTCTCTGCTTAAACC 63

RESULT 5  
US-08-998-416-960  
Sequence 960, Application US/08998416  
Patent No. 6239264  
GENERAL INFORMATION:  
APPLICANT: Philippsen, Peter  
APPLICANT: Pohlmann, Rainer  
APPLICANT: Steiner, Sabine  
APPLICANT: Mohr, Christine  
APPLICANT: Wendland, Jurgen  
APPLICANT: Knechtle, Philipp  
APPLICANT: Reibschung, Corinne  
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPTII  
TITLE OF INVENTION: AND USES THEREOF  
NUMBER OF SEQUENCES: 1152  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: No. 6239264artis Corporation  
STREET: 3054 Cornwallis Road  
CITY: Research Triangle Park  
STATE: No. 6239264ch Carolina  
COUNTRY: USA  
ZIP: 27709  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/998,416  
FILING DATE: 24-DEC-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: CH 0016/97  
FILING DATE: 31-DEC-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Meigs, J. Timothy  
REGISTRATION NUMBER: 38,241  
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8587  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 960:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 588 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: PLAG1592UP  
US-08-998-416-960

Query Match 73.6%; Score 16.2; DB 3; Length 588;  
Best Local Similarity 85.7%; Pred. No. 59;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 ATCTGTACTGCTTAAACC 22  
DB 157 ATCATGTACTGCTTAAACC 177

RESULT 6  
US-09-815-923-1  
Sequence 1, Application US/09815923  
Patent No. 6787642  
GENERAL INFORMATION:

```

APPLICANT: Gill, Sarjeet S.
APPLICANT: Rose, Linda S.
APPLICANT: The Regents of the University of California
TITLE OF INVENTION: Use of Insect Cell Membrane Transporters as No. 6787642e1
FILE REFERENCE: 023070-093800US
CURRENT APPLICATION NUMBER: US/09/815,923
CURRENT FILING DATE: 2001-03-23
NUMBER OF SEQ ID NOS: 20
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 5554
TYPE: DNA
ORGANISM: Manduca sexta
FEATURES:
OTHER INFORMATION: vesicular acetylcholine transporter
US-09-815-923-1

Query Match      73.6%; Score 16.2; DB 4; Length 5554;
Best Local Similarity 85.7%; Pred. No. 82;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 GATCTGTACTGCTTAACC 21
Db      2219 GATCTGTACTGCTTAACC 2239

RESULT 7
US-09-543-681A-4040
Sequence 4040, Application US/09543681A
Patent No. 6605709
GENERAL INFORMATION:
APPLICANT: GARY BRETON
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
FILE REFERENCE: 2709.1002-001
CURRENT APPLICATION NUMBER: US/09/543,681A
CURRENT FILING DATE: 2000-04-05
PRIOR APPLICATION NUMBER: US 60/128,706
PRIOR FILING DATE: 1999-04-09
NUMBER OF SEQ ID NOS: 8344
SEQ ID NO 4040
LENGTH: 741
TYPE: DNA
ORGANISM: Proteus mirabilis
US-09-543-681A-4040

Query Match      72.7%; Score 16; DB 4; Length 741;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      7 GACTGTCTTAACC 22
Db      356 GACTGTCTTAACC 371

RESULT 8
US-09-107-532A-618/C
Sequence 618, Application US/09107532A
Patent No. 6583275
GENERAL INFORMATION:
APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
COMPUTER READABLE FORM:
```

```

MEDIUM TYPE: CD/ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Deneka
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 618:
SEQUENCE CHARACTERISTICS:
LENGTH: 567 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc.feature
LOCATION: (B) LOCATION 1...567
SEQUENCE DESCRIPTION: SEQ ID NO: 618:
US-09-107-532A-618

Query Match      71.8%; Score 15.8; DB 4; Length 567;
Best Local Similarity 89.5%; Pred. No. 93;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      3 TCTTGTACTGCTTAACC 21
Db      474 TCTTGTACTGCTTAACC 456

RESULT 9
US-09-513-999C-10254/C
Sequence 10254, Application US/09513999C
Patent No. 6783961
GENERAL INFORMATION:
APPLICANT: Dumas Milne Edwards, J.B.
APPLICANT: Duclert, A.
APPLICANT: Giordano, J.Y.
TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
FILE REFERENCE: 59.US2.REG
CURRENT APPLICATION NUMBER: US/09/513,999C
CURRENT FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/122,487
PRIOR FILING DATE: 1999-02-26
NUMBER OF SEQ ID NOS: 36681
SOFTWARE: Patent.pm
SEQ ID NO 10254
LENGTH: 346
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc.feature
LOCATION: 125
OTHER INFORMATION: n=a, g, c or t
US-09-513-999C-10254

Query Match      70.9%; Score 15.6; DB 4; Length 346;
```

Best Local Similarity 81.8%; Pred. No. 1.1e+02;  
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GATCTGTACTGCTTAACCC 22  
Db 207 GATCTGAACAGCTCCACCC 186

## RESULT 10

US-09-566-921-2/c  
Sequence 2, Application US/09566921  
Patent No. 668288  
GENERAL INFORMATION:  
APPLICANT: Loring, Jeanne F.  
APPLICANT: Tingley, Debora W.  
APPLICANT: Edwards, Carla M.  
TITLE OF INVENTION: GENES EXPRESSED IN ALZHEIMER'S DISEASE  
FILE REFERENCE: PA-0024 US  
CURRENT APPLICATION NUMBER: US/09/566,921  
CURRENT FILING DATE: 2000-05-05  
NUMBER OF SEQ ID NOS: 138  
SOFTWARE: PERL Program  
SEQ ID NO 2  
LENGTH: 1039  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
OTHER INFORMATION: incyte ID No. 668288 998433.2  
US-09-566-921-2

Query Match 70.9%; Score 15.6; DB 4; Length 1039;  
Best Local Similarity 81.8%; Pred. No. 1.3e+02;  
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GATCTGTACTGCTTAACCC 22  
Db 648 GATCTGAACAGCTCCACCC 627

## RESULT 11

US-09-543-681A-576/c  
Sequence 576, Application US/09543681A  
Patent No. 6605709  
GENERAL INFORMATION:  
APPLICANT: GARY BRETON  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS  
FILE REFERENCE: 2709.1002-001  
CURRENT APPLICATION NUMBER: US/09/543,681A  
CURRENT FILING DATE: 2000-04-05  
PRIOR APPLICATION NUMBER: US 60/128,706  
PRIOR FILING DATE: 1999-04-09  
NUMBER OF SEQ ID NOS: 8344  
SEQ ID NO 576  
LENGTH: 1980  
TYPE: DNA  
ORGANISM: Proteus mirabilis  
US-09-543-681A-576

Query Match 70.9%; Score 15.6; DB 4; Length 1980;  
Best Local Similarity 81.8%; Pred. No. 1.4e+02;  
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GATCTGTACTGCTTAACCC 22  
Db 280 GATCTGTATGCTTAATCC 259

## RESULT 12

US-10-164-595-25  
Sequence 25, Application US/10164595  
Patent No. 6657054

## GENERAL INFORMATION:

APPLICANT: Origene Technologies, Inc  
TITLE OF INVENTION: Regulated Angiogenesis Genes and Polypeptides  
FILE REFERENCE: IU 103 R1  
CURRENT APPLICATION NUMBER: US/10/164,595  
CURRENT FILING DATE: 2002-06-10  
NUMBER OF SEQ ID NOS: 80  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 25  
LENGTH: 1251  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (60)..(1136)  
OTHER INFORMATION:  
US-10-164-595-25

Query Match 70.0%; Score 15.4; DB 4; Length 1251;  
Best Local Similarity 94.1%; Pred. No. 1.7e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 ATCTGTACTGCTTA 18  
Db 1167 ATCTGTACTGCTTA 1183

## RESULT 13

US-08-956-171E-448  
Sequence 448, Application US/08956171E  
Patent No. 6593114  
GENERAL INFORMATION:

APPLICANT: Charles Kunesh

Gil H. Choi

Patrick S. Dillon

Craig A. Rosen

Steven C. Barash

Michael R. Fannon

TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences

NUMBER OF SEQUENCES: 5256

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/956,171E

FILING DATE: 20-Oct-1997

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/009,861

FILING DATE: January 5, 1996

APPLICATION NUMBER: 08/781,986

FILING DATE: January 3, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Mark J. Hyman

REGISTRATION NUMBER: 46,789

REFERENCE/DOCKET NUMBER: PB248P1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (240) 314-1224

TELEFAX: (301) 309-8439

INFORMATION FOR SEQ ID NO: 448:

SEQUENCE CHARACTERISTICS:

LENGTH: 1341 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

```

;
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 448:
US-08-956-171E-448

Query Match      70.0%; Score 15.4; DB 4; Length 1341;
Best Local Similarity 94.1%; Pred. No. 1.7e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 ATCTGTACTGTCCTAA 18
        |||||||
Db      1220 ATCTGTACTGTCCTAA 1236

RESULT 14
US-08-781-986A-448
; Sequence 448, Application US/08781986A
; Patent No. 6737248
; GENERAL INFORMATION:
; APPLICANT: Charles Kunach
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5255
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/781,986A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Benson, Bob
; REGISTRATION NUMBER: 30,446
; REFERENCE/DOCKET NUMBER: PB248PP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 448:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1341 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-781-986A-448

Query Match      70.0%; Score 15.4; DB 4; Length 1341;
Best Local Similarity 94.1%; Pred. No. 1.7e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 ATCTGTACTGTCCTAA 18
        |||||||
Db      1220 ATCTGTACTGTCCTAA 1236

RESULT 15
US-09-220-132-30/c
; Sequence 30, Application US/09220132
; Patent No. 6506607
; GENERAL INFORMATION:
; APPLICANT: Shyjan, Andrew W.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE IDENTIFICATION AND ASSESSMENT
; OF PROSTATE CANCER THERAPIES AND THE DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 07334-074001
```

```

;
; CURRENT APPLICATION NUMBER: US/09/220,132
; CURRENT FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: US 60/079,303
; PRIOR FILING DATE: 1998-03-25
; PRIOR APPLICATION NUMBER: US 60/068,821
; PRIOR FILING DATE: 1997-12-24
; NUMBER OF SEQ ID NOS: 191
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 14770
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-220-132-30

Query Match      70.0%; Score 15.4; DB 4; Length 14770;
Best Local Similarity 94.1%; Pred. No. 2.4e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      3 TCTGTACTGTCCTAA 19
        |||||||
Db      9572 TCTGTACTGTCCTAA 9556
```

Search completed: February 8, 2005, 22:25:57  
Job time : 13.0529 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using bw model

Run on: February 8, 2005, 16:37:21 ; Search time 29.6449 Seconds  
(without alignments)  
4264.119 Million cell updates/sec

Title: US-10-754-437-5

Perfect score: 22  
Sequence: 1 gatctgtactgtcttaaaccc 22

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4300275 seqs, 2872944193 residues

Total number of hits satisfying chosen parameters: 8600550

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:\*

- 1: /cgn2\_6/prodata/2/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/prodata/2/pubpna/US06\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/prodata/2/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/prodata/2/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/prodata/2/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/prodata/2/pubpna/US07\_PUBCOMB.seq:\*
- 7: /cgn2\_6/prodata/2/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/prodata/2/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/prodata/2/pubpna/US09A\_PUBCOMB.seq:\*
- 10: /cgn2\_6/prodata/2/pubpna/US09B\_PUBCOMB.seq:\*
- 11: /cgn2\_6/prodata/2/pubpna/US09C\_PUBCOMB.seq:\*
- 12: /cgn2\_6/prodata/2/pubpna/US09\_NEW\_PUB.seq:\*
- 13: /cgn2\_6/prodata/2/pubpna/US10\_PUBCOMB.seq:\*
- 14: /cgn2\_6/prodata/2/pubpna/US10B\_PUBCOMB.seq:\*
- 15: /cgn2\_6/prodata/2/pubpna/US10C\_PUBCOMB.seq:\*
- 16: /cgn2\_6/prodata/2/pubpna/US10D\_PUBCOMB.seq:\*
- 17: /cgn2\_6/prodata/2/pubpna/US10\_PUBCOMB.seq:\*
- 18: /cgn2\_6/prodata/2/pubpna/US10\_NEW\_PUB.seq:\*
- 19: /cgn2\_6/prodata/2/pubpna/US11\_NEW\_PUB.seq:\*
- 20: /cgn2\_6/prodata/2/pubpna/US60\_NEW\_PUB.seq:\*
- 21: /cgn2\_6/prodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	17.2	78.2	220	18	US-10-674-124A-21842
2	17.2	78.2	454	10	US-09-918-995-9955
3	17.2	78.2	470	9	US-09-770-444-201
4	17.2	78.2	600	13	US-10-027-632-287049
5	17.2	78.2	600	13	US-10-027-632-287050
6	17.2	78.2	600	15	US-10-027-632-287049
7	17.2	78.2	600	15	US-10-027-632-287050
8	17.2	78.2	3355	13	US-10-027-632-114078
9	17.2	78.2	3355	15	US-10-027-632-114078
10	17.2	78.2	39079	17	US-10-322-281-832
11	17.2	78.2	101569	17	US-10-717-597-10
12	17.2	78.2	259202	18	US-10-723-860-1311

13	17	77.3	77834	13	US-10-087-192-343	Sequence 343, App
14	16.8	76.4	567	11	US-09-864-408A-7139	Sequence 1719, App
15	16.8	76.4	568	18	US-10-425-115-149520	Sequence 149520,
16	16.8	76.4	835	17	US-10-437-963-93764	Sequence 93764, A
17	16.8	76.4	1301	17	US-10-767-701-9447	Sequence 9447, App
18	16.8	76.4	2568	17	US-10-437-963-93762	Sequence 93762, A
19	16.4	74.5	508	9	US-09-864-761-12341	Sequence 12341, A
20	16.4	74.5	593	13	US-10-027-632-47128	Sequence 47128, A
21	16.4	74.5	593	15	US-10-027-632-47128	Sequence 47128, A
22	16.4	74.5	625	16	US-10-424-599-117287	Sequence 117287, A
23	16.4	74.5	777	13	US-10-027-632-21575	Sequence 21575, A
24	16.4	74.5	777	15	US-10-027-632-21575	Sequence 21575, A
25	16.4	74.5	139257	10	US-09-920-671-11	Sequence 11, App1
26	16.2	73.6	281	16	US-10-424-599-68086	Sequence 68086, A
27	16.2	73.6	339	17	US-10-469-285-288	Sequence 288, App
28	16.2	73.6	1039	17	US-10-767-701-15302	Sequence 15302, A
29	16.2	73.6	1661	18	US-10-739-930-5539	Sequence 5539, App
30	16.2	73.6	1760	17	US-10-767-701-15450	Sequence 15450, A
31	16.2	73.6	1777	16	US-10-425-114-31265	Sequence 31265, A
32	16.2	73.6	1787	16	US-10-425-114-24165	Sequence 24165, A
33	16.2	73.6	2000	9	US-09-938-842A-4001	Sequence 4001, App
34	16.2	73.6	2000	11	US-09-938-842A-4001	Sequence 4001, App
35	16.2	73.6	2515	18	US-10-425-115-58255	Sequence 58255, A
36	16.2	73.6	3239	13	US-10-027-632-114978	Sequence 114978,
37	16.2	73.6	3239	15	US-10-027-632-114978	Sequence 114978,
38	16.2	73.6	5554	9	US-09-815-923-1	Sequence 1, App1
39	16.2	73.6	14759	10	US-09-764-891-8344	Sequence 8344, App
40	16.2	73.6	14759	15	US-10-205-428-777	Sequence 777, App
41	16.2	73.6	17146	9	US-09-764-877-3850	Sequence 3850, App
42	16.2	73.6	17146	16	US-10-242-515-3850	Sequence 3850, App
43	16.2	73.6	30568	9	US-09-764-877-3851	Sequence 3851, App
44	16.2	73.6	30568	16	US-10-242-515-3851	Sequence 3851, App
45	16.2	73.6	49031	17	US-10-322-281-523	Sequence 523, App

#### ALIGNMENTS

RESULT 1

US-10-674-124A-21842

Sequence 21842, Application US/10674124A

Publication No. US20040197797A1

GENERAL INFORMATION:

APPLICANT: INOKO, Hidetoshi

APPLICANT: TAMITA, Gen

TITLE OF INVENTION: GENE MAPPING METHOD USING MICROSATELLITE

FILE REFERENCE: ORIN-003CIP

CURRENT APPLICATION NUMBER: US/10/674,124A

CURRENT FILING DATE: 2003-09-26

PRIOR APPLICATION NUMBER: 10/257,511

PRIOR FILING DATE: 2003-03-07

PRIOR APPLICATION NUMBER: PCT/JP00/07621

PRIOR FILING DATE: 2000-10-30

PRIOR APPLICATION NUMBER: JP2000-112699

PRIOR FILING DATE: 2000-04-13

PRIOR APPLICATION NUMBER: JP2002-327516

PRIOR FILING DATE: 2002-09-28

PRIOR APPLICATION NUMBER: JP2002-383869

PRIOR FILING DATE: 2002-12-09

NUMBER OF SEQ ID NOS: 27110

SEQ ID NO 21842

LENGTH: 220

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

OTHER INFORMATION: Z66933

FEATURE:

OTHER INFORMATION: Located on chromosome 16

OTHER INFORMATION: Distance between a terminus base of telomere on

OTHER INFORMATION: chromosomal short arm and 5'-terminus of this base

OTHER INFORMATION: sequence : 9358990

```
FEATURE:
; OTHER INFORMATION: Distance between 3'-terminus of neighbour sequence of
; OTHER INFORMATION: sequence listing upward to telomere on chromosomal short arm and
; OTHER INFORMATION: 5'-terminus of this base sequence : 5631
US-10-674-124A-21842

Query Match      78.2%; Score 17.2; DB 18; Length 220;
Best Local Similarity 86.4%; Pred. No. 1.9e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 GATCTGTACTGTCTTAACCC 22
Db      148 GATCTGTAAATGTTGAACCC 169

RESULT 2
US-09-918-995-9955/c
; Sequence 9955, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-736
; CURRENT APPLICATION NUMBER: US/09/918,995
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 9955
; LENGTH: 454
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-918-995-9955

Query Match      78.2%; Score 17.2; DB 10; Length 454;
Best Local Similarity 86.4%; Pred. No. 2.2e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 GATCTGTACTGTCTTAACCC 22
Db      140 GATCTGTCTGTCTTAACCTC 119

RESULT 3
US-09-770-444-201
; Sequence 201, Application US/09770444
; Patent No. US20020023280A1
; GENERAL INFORMATION:
; APPLICANT: Gorlach, Jorn
; APPLICANT: An, Yong-Qiang
; APPLICANT: Hamilton, Carol M.
; APPLICANT: Price, Jennifer L.
; APPLICANT: Raines, Tracy M.
; APPLICANT: Yu, Yang
; APPLICANT: Rameaka, Joshua G.
; APPLICANT: Page, Amy
; APPLICANT: Matthew, Abraham V.
; APPLICANT: Ledford, Brooke L.
; APPLICANT: Moesener, Jeffrey P.
; APPLICANT: Garcia, Carlos A.
; APPLICANT: Krickler, Maja
; APPLICANT: Slader, Ted
; APPLICANT: Davis, Keith R.
; APPLICANT: Allen, Keith
; APPLICANT: Hoffman, Neil
; APPLICANT: Hurban, Patrick
; TITLE OF INVENTION: Expressed Sequences of Arabidopsis
; FILE REFERENCE: 2027 (PAPA-016PRV)
; CURRENT APPLICATION NUMBER: US/09/770,444
```

```
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: 60/178,502
; PRIOR FILING DATE: 2000-01-27
; NUMBER OF SEQ ID NOS: 999
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 201
; LENGTH: 470
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-770-444-201

Query Match      78.2%; Score 17.2; DB 9; Length 470;
Best Local Similarity 86.4%; Pred. No. 2.2e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 GATCTGTACTGTCTTAACCC 22
Db      292 GATCTGTCTTGTCTTAACCC 313

RESULT 4
US-10-027-632-287049
; Sequence 287049, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 287049
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-287049

Query Match      78.2%; Score 17.2; DB 13; Length 600;
Best Local Similarity 86.4%; Pred. No. 2.3e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 GATCTGTACTGTCTTAACCC 22
Db      99 GATCTCACTGTTTAAACCC 120

RESULT 5
US-10-027-632-287050
; Sequence 287050, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
```



```

; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 287050
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-287050
```

```
Query Match      78.2% Score 17.2; DB 13; Length 600;
Best Local Similarity 86.4%; Pred. No. 2.3e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY      1 GATCTGTACTGCTTTAAACC 22
      ||||| ||||| ||||| |||||
DB      99 GATCTCATCTGTTTAAACC 120
```

```

RESULT 6
US-10-027-632-287049
; Sequence 287049, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 287049
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-287049
```

```
Query Match      78.2% Score 17.2; DB 15; Length 600;
Best Local Similarity 86.4%; Pred. No. 2.3e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY      1 GATCTGTACTGCTTTAAACC 22
      ||||| ||||| ||||| |||||
DB      99 GATCTCATCTGTTTAAACC 120
```

```

RESULT 7
US-10-027-632-287050
; Sequence 287050, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 287050
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-287050
```

```
Query Match      78.2% Score 17.2; DB 15; Length 600;
Best Local Similarity 86.4%; Pred. No. 2.3e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY      1 GATCTGTACTGCTTTAAACC 22
      ||||| ||||| ||||| |||||
DB      99 GATCTCATCTGTTTAAACC 120
```

```

RESULT 8
US-10-027-632-114078
; Sequence 114078, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 114078
; LENGTH: 3355
; TYPE: DNA
; ORGANISM: Human
```

US-10-027-632-114078

Query Match 78.2% Score 17.2; DB 13; Length 3355;  
Best Local Similarity 86.4%; Pred. No. 3e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GATCTGTACTGCTTAACCC 22  
Db 2696 GATCTGTACTGCTTAATCCC 2717

RESULT 9

US-10-027-632-114078  
; Sequence 114078, Application US/10027632  
; Publication No. US20030204075A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; TITLE OF INVENTION: Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/228,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 114078  
; LENGTH: 3355  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-114078

Query Match 78.2% Score 17.2; DB 15; Length 3355;  
Best Local Similarity 86.4%; Pred. No. 3e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GATCTGTACTGCTTAACCC 22  
Db 2696 GATCTGTACTGCTTAATCCC 2717

RESULT 10

US-10-322-281-832/c  
; Sequence 832, Application US/10322281  
; Publication No. US20040126762A1  
; GENERAL INFORMATION:  
; APPLICANT: David W. Morris  
; APPLICANT: Marc S. Malandro  
; TITLE OF INVENTION: Novel Compositions and Methods in Cancer  
; FILE REFERENCE: 529452001000  
; CURRENT APPLICATION NUMBER: US/10/322,281  
; CURRENT FILING DATE: 2002-12-17  
; NUMBER OF SEQ ID NOS: 866  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 832  
; LENGTH: 39079  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-322-281-832

Query Match 78.2% Score 17.2; DB 17; Length 39079;

Best Local Similarity 86.4%; Pred. No. 4.6e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GATCTGTACTGCTTAACCC 22  
Db 37200 GATCTGAACCTGCTTAATCCC 37179

RESULT 11

US-10-717-597-10/c  
; Sequence 10, Application US/10717597  
; Publication No. US20040110221A1  
; GENERAL INFORMATION:  
; APPLICANT: Wyeth  
; APPLICANT: Burczynski, Michael E.  
; APPLICANT: Twine, Natalie C.  
; APPLICANT: Dornier, Andrew J.  
; APPLICANT: Trepicchio, William L.  
; APPLICANT: Stonim, Donna K.  
; APPLICANT: Stover, Jennifer A.  
; TITLE OF INVENTION: METHODS FOR DIAGNOSING RCC AND OTHER SOLID TUMORS  
; FILE REFERENCE: AM101080L  
; CURRENT APPLICATION NUMBER: US/10/717,597  
; CURRENT FILING DATE: 2003-11-21  
; PRIOR APPLICATION NUMBER: US 60/459,782  
; PRIOR FILING DATE: 2003-04-03  
; PRIOR APPLICATION NUMBER: US 60/427,982  
; PRIOR FILING DATE: 2002-11-21  
; NUMBER OF SEQ ID NOS: 4904  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 10  
; LENGTH: 101569  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (22955)..(22955)  
; OTHER INFORMATION: No residue exists at "n" position (position 22955) in the  
; OTHER INFORMATION: corresponding human genome sequence (ETS2 gene) in the Entrez  
; OTHER INFORMATION: Human Genome Sequence Database.  
US-10-717-597-10

Query Match 78.2% Score 17.2; DB 17; Length 101569;  
Best Local Similarity 86.4%; Pred. No. 5.4e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GATCTGTACTGCTTAACCC 22  
Db 55369 GATCTGAACCTGCTTAATCCC 55348

RESULT 12

US-10-723-860-1311/c  
; Sequence 1311, Application US/10723860  
; Publication No. US20040253606A1  
; GENERAL INFORMATION:  
; APPLICANT: Aziz, Natascha  
; APPLICANT: Ginsburg, Wendy M.  
; APPLICANT: Zlotnik, Albert  
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &  
; TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators  
; FILE REFERENCE: 05882.0193.NPUS01  
; CURRENT APPLICATION NUMBER: US/10/723,860  
; CURRENT FILING DATE: 2003-11-26  
; PRIOR APPLICATION NUMBER: 60/429,739  
; PRIOR FILING DATE: 2002-11-26  
; NUMBER OF SEQ ID NOS: 8393  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 1311  
; LENGTH: 259202  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-723-860-1311

Query Match 78.2%; Score 17.2; DB 18; Length 259202;  
Best Local Similarity 86.4%; Pred. No. 6.2e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GATCTGTACTGTCTTAAACC 22  
DB 70030 GATCTCATCTGTCTTAAACC 70009

RESULT 13  
US-10-087-192-343/C  
Sequence 343, Application US/10087192  
Publication No. US20020182586A1  
GENERAL INFORMATION:  
APPLICANT: Morris, David W.  
APPLICANT: Engelhard, Eric K.  
TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR  
TITLE OF INVENTION: CANCER  
FILE REFERENCE: 529452000122  
CURRENT APPLICATION NUMBER: US/10/087,192  
CURRENT FILING DATE: 2002-03-01  
PRIOR APPLICATION NUMBER: US 09/747,377  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: US 09/798,586  
PRIOR FILING DATE: 2001-03-02  
NUMBER OF SEQ ID NOS: 2059  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 343  
LENGTH: 77834  
TYPE: DNA  
ORGANISM: Mus musculus  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(77834)  
OTHER INFORMATION: n = A,T,C or G  
US-10-087-192-343

Query Match 77.3%; Score 17; DB 13; Length 77834;  
Best Local Similarity 100.0%; Pred. No. 6.4e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 TCTTGACTGTCTTAA 19  
DB 64684 TCTTGACTGTCTTAA 64668

RESULT 14  
US-09-864-408A-7139/C  
Sequence 7139, Application US/09864408A  
Publication No. US2004009474A1  
GENERAL INFORMATION:  
APPLICANT: Leach, Martin D.  
APPLICANT: Shinkets, Richard A.  
TITLE OF INVENTION: No. US2004009474A1 Human Polynucleotides and Polypeptides Enc  
FILE REFERENCE: 21402-012  
CURRENT APPLICATION NUMBER: US/09/864,408A  
CURRENT FILING DATE: 2001-05-24  
PRIOR APPLICATION NUMBER: 60/206,690  
PRIOR FILING DATE: 2000-05-24  
NUMBER OF SEQ ID NOS: 9068  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 7139  
LENGTH: 567  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-864-408A-7139

Query Match 76.4%; Score 16.8; DB 11; Length 567;  
Best Local Similarity 90.0%; Pred. No. 3.5e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 TCTTGACTGTCTTAAACC 22

DB 303 TCTTGACTGTCTTAAACC 284

RESULT 15  
US-10-425-115-149520  
Sequence 149520, Application US/10425115  
Publication No. US20040214272A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa, Thomas J.  
APPLICANT: Kovalic, David K.  
APPLICANT: Zhou, Yihua  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
TITLE OF INVENTION: Plants  
FILE REFERENCE: 38-21(53222)B  
CURRENT APPLICATION NUMBER: US/10/425,115  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 369326  
SEQ ID NO 149520  
LENGTH: 568  
TYPE: DNA  
ORGANISM: Zea mays  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (1)..(568)  
OTHER INFORMATION: unsure at all n locations  
FEATURE:  
OTHER INFORMATION: Clone ID: MRT4577\_6788C.1  
US-10-425-115-149520

Query Match 76.4%; Score 16.8; DB 18; Length 568;  
Best Local Similarity 90.0%; Pred. No. 3.5e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 TCTTGACTGTCTTAAACC 22  
DB 36 TCTTGACTGTCTTAAACC 55

Search completed: February 9, 2005, 06:50:56  
Job time : 33.6449 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 8, 2005, 12:58:25 ; Search time 262.041 Seconds  
(without alignments)  
3059.349 Million cell updates/sec

Title: US-10-754-437-5

Perfect score: 22  
Sequence: 1 gatcttgactgcttcaacc 22

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

EST:\*  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_est9:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	18.8	85.5	300	1	AJ461766
C 2	18.8	85.5	423	1	AV692067
C 3	17.8	80.9	505	1	AU003892
C 4	17.8	80.9	606	8	B2419788
C 5	17.8	80.9	711	9	AG402501
C 6	17.8	80.9	799	4	CG638234
C 7	17.8	80.9	804	4	CG289973
C 8	17.8	80.9	895	9	CG320249
C 9	17.8	80.9	932	9	CG324260
C 10	17.4	79.1	500	5	BP185852
C 11	17.4	79.1	685	9	CE153858
C 12	17.4	79.1	700	9	CG922134
C 13	17.4	79.1	740	9	AG452989
C 14	17.2	78.2	242	6	CB261729
C 15	17.2	78.2	300	1	AU099200
C 16	17.2	78.2	310	1	AA380417
C 17	17.2	78.2	310	1	AA227042
C 18	17.2	78.2	321	6	CD529079
C 19	17.2	78.2	346	8	AZ510639
C 20	17.2	78.2	360	6	CS1188
C 21	17.2	78.2	360	7	D67108
C 22	17.2	78.2	369	5	BP654637
C 23	17.2	78.2	376	1	A1870075
C 24	17.2	78.2	376	5	BP649093

25	17.2	78.2	384	8	A0170514
26	17.2	78.2	388	1	AV799434
27	17.2	78.2	398	4	BM755387
28	17.2	78.2	402	1	AV789942
29	17.2	78.2	402	3	CK120300
30	17.2	78.2	403	5	BP647245
31	17.2	78.2	406	1	AV815144
32	17.2	78.2	411	1	AA597625
33	17.2	78.2	418	1	AV788538
34	17.2	78.2	421	6	CB263536
35	17.2	78.2	423	7	CK118559
36	17.2	78.2	427	7	R90061
37	17.2	78.2	431	5	BP626409
38	17.2	78.2	432	3	CN50A6EL
39	17.2	78.2	434	3	CN50A6EL
40	17.2	78.2	437	6	CB265220
41	17.2	78.2	438	7	CK121985
42	17.2	78.2	440	8	AQ050661
43	17.2	78.2	441	7	CK118716
44	17.2	78.2	444	7	CK117689
45	17.2	78.2	444	7	CK119024

## ALIGNMENTS

RESULT 1  
AJ461766/c 300 bp mRNA linear EST 24-MAY-2002  
DEFINITION  
S0000200021A12P1, mRNA sequence.

ACCESSION  
AJ461766  
VERSION  
AJ461766.1 GI:21060686  
KEYWORDS  
SOURCE  
ORGANISM

REFERENCE  
Saren, A.-M., Tanekainen, J., Paulin, L. and Schulman, A.H.  
TITLE  
Barley EST's  
JOURNAL  
Unpublished (2002)  
COMMENT  
Contact: Schulman AH  
Institute of Biotechnology  
University of Helsinki  
P.O. Box 56 (Viikinkaari 6A), University of Helsinki FIN-00014, Finland.

FEATURES  
source  
1..300  
Location/Qualifiers  
/organism="Hordeum vulgare subsp. vulgare"  
/mol\_type="mRNA"  
/cultiivar="Saana"  
/sub\_species="vulgare"  
/db\_xref="taxon:112509"  
/dev\_stage="Embryo"  
/clone\_11b="S00002"  
/note="1 day after pollination"

ORIGIN  
Query Match 85.5%; Score 18.8; DB 1; Length 300;  
Best Local Similarity 90.9%; Pred. No. 1.9e+02;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GATCTTGACTGCTTCAACC 22  
|||||||  
DB 212 GATCTTGACTGCTTCAACC 191  
|||||||

RESULT 2  
AV692067/c 423 bp mRNA linear EST 16-JAN-2002

```

FEATURES
  source
    Location/Qualifiers
      1..505
        /organism="Bombyx mori"
        /mol_type="mRNA"
        /strain="p50 (Daizo)"
        /db_xref="taxon:7091"
        /clone="wgs0710"
        /clone_id="Bombyx mori p50 (Daizo)"

ORIGIN
Query Match      80.9%; Score 17.8; DB 1; Length 505;
Best Local Similarity 90.5%; Pred. No. 6,3e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 GATCTTGACTGCTTAACC 21
        |||||||
        5 GTCTTGACTGGCTTAACC 25

LOCUS      BZ419788      606 bp      DNA      linear      GSS 10-DEC-2002
DEFINITION BZ419788.1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
            1f59f06 5', genomic survey sequence.
ACCESSION  BZ419788
VERSION     BZ419788.1
KEYWORDS    GS.
SOURCE      Zea mays
            Zea mays
            Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoidae; Andropogoneae; Zea.
            1 (bases 1 to 606)
REFERENCE   Rabinowitz,P.D., O'Shaughnessy,A.L., Balija,V., Dedhia,N.,
            Kaczendzinger,F., King,L., Miller,B., Muller,S., Nascimento,L.,
            Zultewern,T., McCombie,W.R. and Martienssen,R.A.
            Genomic shotgun sequences from Zea mays (methyl-filtered)
            Unpublished (2002)
            Contact: W. Richard McCombie
            Iltis Annenberg Hazen Genome Sequencing Center
            Cold Spring Harbor Laboratory
            PO Box 100, Cold Spring Harbor, NY 11724, USA
            Tel: 516 367 8884
            Fax: 516 367 8874
            Email: mcombie@cehl.org
            Plate: 1f59 row: f column: 06
            Seq primer: -21M13univFwd
            Class: shotgun
            High quality sequence stop: 606.
FEATURES
  source
    Location/Qualifiers
      1..606
        /organism="Zea mays"
        /mol_type="genomic DNA"
        /cultivar="B73"
        /db_xref="taxon:4577"
        /clone="1f59f06"
        /lab_host="DH5a"
        /clone_id="WGS-ZmaysF (DH5a methyl filtered)"
        /notes="Organ: immature ears; Site_1: Xba I; Site_2: Xba I;
        The vector was digested with XbaI and one nucleotide was
        added by fill in in the recessive 3' end. The genomic DNA
        was nebulized, end repaired, adaptor ligated and size
        fractionated using sephadex. The resulting fragments were
        between 0.8 and 3 kb and were cloned into the vector (x/y
        reads in M13mp19, b/g reads in pUC19). The same ligation
        was transformed into DH5a."

ORIGIN
Query Match      80.9%; Score 17.8; DB 8; Length 606;
Best Local Similarity 90.5%; Pred. No. 6,4e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

Db      15 ATCTTGTCGCTTAAACC 35
|||||
RESULT 5
LOCUS   AG402501
DEFINITION Mus musculus molossinus DNA, clone:MSMg01-237J13.T7, genomic survey
SEQUENCE
ACCESSION AG402501 GI:48045187
KEYWORDS GSS.
SOURCE    Mus musculus molossinus
ORGANISM  Mus musculus molossinus
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
TITLE    BAC end Sequences of Library MSMg01
JOURNAL  Unpublished
REFERENCE 2 (bases 1 to 711)
AUTHORS  Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
TITLE    Direct Submission
JOURNAL  Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
          and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
          1-7-22 Suenho-chou,Tsukumi-Ku, Yokohama, Kanagawa 230-0045, Japan
          E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
          Tel:81-45-503-9111, Fax:81-45-503-9170)
          Clones are derived from the mouse BAC library MSMg01. For BAC
          library availability, please contact Kunya Abe (abe@erc.riken.jp).
          Tsukuba Institute, Bio Resource Center.
          The Institute of Physical and Chemical Research (RIKEN) 3-1-1
          Koyada, Tsukuba, 305-0074 Japan
          phone: 81-298-36-9189, fax: 81-298-36-9199
          e-mail: abe@erc.riken.jp
          PRIMERS
          Sequencing : T7
          LIBRARY
          Vector : pBACe3.6
          R.Site 1 : EcoRI
          R.Site 2 : EcoRI.
FEATURES
source    Location/Qualifiers
          1..711
             /organism="Mus musculus molossinus"
             /mol_type="genomic DNA"
             /sub_species="molossinus"
             /db_xref="taxon:57486"
             /clone="MSMg01-237J13.T7"
             /sex="male"
             /tissue_type="mixture of kidney and spleen"
             /clone_lib="MSMg01 Mouse Male BAC Library"

ORIGIN
Query Match      80.9%; Score 17.8; DB 9; Length 711;
Best Local Similarity 90.5%; Pred. No. 6.5e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 ATCTTGTCGCTTAAACC 22
|||||
Db 350 ATCTTGTCGCTTAAACC 370
|||||

RESULT 6
LOCUS   CC638234
DEFINITION OGVBF41TH ZM 0.7.1.5 KB Zea mays genomic clone ZMBWA0498G10,
          genomic survey sequence.
ACCESSION CC638234 GI:32019113
KEYWORDS GSS.
SOURCE    Zea mays
ORGANISM  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

```

```

          Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
          clade; Panicoideae; Andropogoneae; Zea.
REFERENCE
AUTHORS WhiteIaw,C.A., Quackenbush,J., Van Aken,S., Uterback,T.,
          Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
          Citek,R.W., Numborg,A., Robbins,D. and Lakey,N.
TITLE    Consortium for Maize Genomics
JOURNAL  Unpublished (2002)
COMMENT  Other_GSSs: OGVBF41TV
          Contact: Cathy WhiteIaw
          TIGR
          712 Medical Center Drive, Rockville, MD 20850, USA
          Tel: 301-838-5843
          Fax: 301-838-0208
          Email: whiteIaw@tigr.org
          Seq primer: TR
          Class: sheared ends.
FEATURES
source    Location/Qualifiers
          1..799
             /organism="Zea mays"
             /mol_type="genomic DNA"
             /strain="B73"
             /db_xref="taxon:4577"
             /clone="ZMBWA0498G10"
             /clone_lib="ZM 0.7.1.5 KB"
             /note="vector: pBCK-, Site_1: HincII; 0.7-1.5 kb
             methylation filtered genomic DNA library"

ORIGIN
Query Match      80.9%; Score 17.8; DB 9; Length 799;
Best Local Similarity 90.5%; Pred. No. 6.6e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 ATCTTGTCGCTTAAACC 22
|||||
Db 67 ATCTTGTCGCTTAAACC 87
|||||

RESULT 7
LOCUS   BG299973
DEFINITION HVSMA0022M13f Hordeum vulgare seedling shoot EST library
          HVCDA0001 (Cold stress) Hordeum vulgare subsp. vulgare cDNA clone
          HVSMA0022M13f, mRNA sequence.
ACCESSION BG299973
KEYWORDS EST.
SOURCE    Hordeum vulgare subsp. vulgare
ORGANISM  Hordeum vulgare subsp. vulgare
          Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
          Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
          Pooidae; Triticeae; Hordeum.
          1 (bases 1 to 804)
          Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Begum,D., Frisch,D.,
          Yu,Y., Henry,D., Palmer,M., Rambo,T., Simmons,J., Oates,R.,
          Choi,D.W., Fenton,R.D. and Main,D.
          Development of a genetically and physically anchored EST resource
          for barley genomics: Morex cold-stressed seedling shoot cDNA
          library
          Unpublished (2001)
          Contact: Wing RA
          Clemson University Genomics Institute
          Clemson University
          100 Jordan Hall, Clemson, SC 29634, USA
          Tel: 864 656 7288
          Fax: 864 656 4293
          Email: rwing@clemson.edu
          Total hg bases = 228
          Seq primer: ATTTAACCCTCAATAAGG
          High quality sequence start: 7
          High quality sequence stop: 464.
          Location/Qualifiers
          1..804

```

```

/organism="Hordeum vulgare subsp. vulgare"
/mol_type="mRNA"
/cultivar="Morex"
/sub_species="vulgare"
/db_xref="taxon:112509"
/clone="HVSME0022M13f"
/issue_type="Seedling shoot"
/lab_host="TUC121"
/clone_lib="Hordeum vulgare seedling shoot EST library
HVCNDA0001 (Cold stress)"
/notes="Vector: lambdaZAP; Site 1: EcoRI; Site 2: XhoI;
Seeds were surface sterilized then germinated under aseptic
conditions in the dark at room temperature on filter paper
with water, nystatin and cefotaxime in covered
crystallization dishes. Five-day old seedlings were
incubated at 50C for 2 days. Shoots were then harvested,
total RNA was prepared, poly(A) RNA was purified, one
primary unamplified cDNA library was made, and 600000 pfu
were in vivo excised to give Bluescript SK(-) cDNA
phagemide. These steps were performed in the TU close
laboratory at the University of California, Riverside
(Choi, Close, Fenton). Phagemids were plated and picked at
the Clemson University Genomics Institute (CUGI) (Begum,
Palmer, Friesch, Atkins and Wang). Plasmid DNA
preparations, DNA sequencing and sequence analysis were
performed at CUGI (Wing, Yu, Friesch, Henry, Simmons,
Oates, Rambo, Main). The sequence has been trimmed to
remove vector sequence and contains a minimum of 100 bases
of phred value 20 or above. For more details on library
preparation and sequence analysis see
http://www.genome.clemson.edu/projects/barley. To order
this clone see http://www.genome.clemson.edu/orders Also
see Close TJ, Wing R, Kleinholz A, Wise R (2001)
Genetically and physically anchored EST resources for
barley genomics. Barley Genetics Newsletter 31:29-30.
(http://wheat.pw.usda.gov/gsgpages/bgn/31/cover.html)"

```

## ORIGIN

```

Query Match      80.9%; Score 17.8; DB 4; Length 804;
Best Local Similarity 90.5%; Pred. No. 6.6e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY      1 GATCTGACTGCTTAACC 21
      1 |||||
Db      483 GTCTTGTACTGTCTGAACC 503

```

## RESULT 8

```

CG320249/c      895 bp      DNA      linear      GSS 26-AUG-2003
LOCUS      OG3BP48TV_ZM_0.7_1.5_KB_Zea_mays genomic clone ZMMBMA0770G24,
DEFINITION      genomic survey sequence.
ACCESSION      CG320249
VERSION      CG320249.1 GI:34237515
KEYWORDS      GSS.
SOURCE      Zea mays
ORGANISM      Zea mays

```

```

REFERENCE
AUTHORS      Whiteclaw,C.A., Quackenbush,J., Van Aken,S., Uterback,T.,
      Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
      Citek,R.W., Numborg,A., Robbins,D. and Lakey,N.
      1 (bases 1 to 895)
      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
      Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
      clade; Panicoideae; Andropogoneae; Zea.

```

```

TITLE      Unpublished (2002)
JOURNAL      Consortium for Maize Genomics
COMMENT      Other GSSs: OG3BP48TH
      Contact: Cathy Whiteclaw
      TIGR
      9712 Medical Center Drive, Rockville, MD 20850, USA
      Tel: 301-838-5843
      Fax: 301-838-0208
      Email: whiteclaw@tigr.org

```

```

FEATURES
      source
      Class: sheared ends.
      Location/Qualifiers
      1..895
      /organism="Zea mays"
      /mol_type="genomic DNA"
      /strain="B73"
      /db_xref="taxon:4577"
      /clone="ZMMBMA0770G24"
      /clone_lib="ZM_0.7_1.5_KB"
      /note="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb
      methylation filtered genomic DNA library"

```

## ORIGIN

```

Query Match      80.9%; Score 17.8; DB 9; Length 895;
Best Local Similarity 90.5%; Pred. No. 6.7e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY      2 ATCTGTACTGCTTAACC 22
      2 |||||
Db      233 ATCTGTCTGCTTAACC 213

```

## RESULT 9

```

CG234260      932 bp      DNA      linear      GSS 22-AUG-2003
LOCUS      OGZAG01TV_ZM_0.7_1.5_KB_Zea_mays genomic clone ZMMBMA0656B01,
DEFINITION      genomic survey sequence.
ACCESSION      CG234260
VERSION      CG234260.1 GI:34134146
KEYWORDS      GSS.
SOURCE      Zea mays
ORGANISM      Zea mays

```

```

REFERENCE
AUTHORS      Whiteclaw,C.A., Quackenbush,J., Van Aken,S., Uterback,T.,
      Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
      Citek,R.W., Numborg,A., Robbins,D. and Lakey,N.
      1 (bases 1 to 932)
      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
      Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
      clade; Panicoideae; Andropogoneae; Zea.

```

```

TITLE      Unpublished (2002)
JOURNAL      Consortium for Maize Genomics
COMMENT      Other GSSs: OGZAG01TH
      Contact: Cathy Whiteclaw
      TIGR
      9712 Medical Center Drive, Rockville, MD 20850, USA
      Tel: 301-838-5843
      Fax: 301-838-0208
      Email: whiteclaw@tigr.org

```

```

FEATURES
      source
      Class: sheared ends.
      Location/Qualifiers
      1..932
      /organism="Zea mays"
      /mol_type="genomic DNA"
      /strain="B73"
      /db_xref="taxon:4577"
      /clone="ZMMBMA0656B01"
      /clone_lib="ZM_0.7_1.5_KB"
      /note="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb
      methylation filtered genomic DNA library"

```

## ORIGIN

```

Query Match      80.9%; Score 17.8; DB 9; Length 932;
Best Local Similarity 90.5%; Pred. No. 6.7e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY      2 ATCTGTACTGCTTAACC 22
      2 |||||
Db      276 ATCTGTCTGCTTAACC 296

```

## RESULT 10



BP185852/c  
LOCUS BP185852 500 bp mRNA linear EST 17-JUN-2003  
DEFINITION BP185852 planarian head cDNA Dugesia japonica cDNA clone 03031\_HH,  
mRNA sequence.  
ACCESSION BP185852  
VERSION BP185852.1 GI:32899764  
KEYWORDS EST.  
SOURCE Dugesia japonica  
ORGANISM Eukaryota; Metazoa; Platyhelminthes; Turbellaria; Sericata;  
Tricladida; Paludicola; Dugesidae; Dugesia.  
1 (bases 1 to 500)  
Mineta, K., Nakazawa, M., Cebria, F., Ikeo, K., Agata, K. and  
Gojobori, T.  
Origin and evolutionary process of the CNS elucidated by  
comparative genomics analysis of planarian ESTs  
Proc. Natl. Acad. Sci. U.S.A. 100 (13), 7666-7671 (2003)  
22709146  
MEDLINE  
PUBMED 12802012  
COMMENT Contact: Katsuniko Mineta  
National Institute of Genetics, Center for Information Biology and  
DNA Data Bank of Japan  
1111 Yata, Mishima, Shizuoka 411-8540, Japan  
Tel: 81-559-81-6847  
Fax: 81-559-81-6848  
Email: kmineta@lab.nig.ac.jp  
These clones and additional information are obtained from our web  
site: <http://www.cdb.nig.ac.jp/dda/>.  
Location/Qualifiers  
1..500  
/organism="Dugesia japonica"  
/mol\_type="mRNA"  
/db\_xref="taxon:6161"  
/clone="03031\_HH"  
/issue\_type="head"  
/dev\_stage="adult"  
/clone\_lib="planarian head cDNA"

ORIGIN  
Query Match 79.1%; Score 17.4; DB 5; Length 500;  
Best Local Similarity 94.7%; Pred. No. 1e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GATCTTGACTGCTTTAA 19  
|||||  
DB 429 GATCTTGACTGCTTCAA 411

RESULT 11  
CE153858/c  
LOCUS CE153858 685 bp DNA linear GSS 25-SEP-2003  
DEFINITION tigr-gss-dog-17000371358796 Dog Library Canis familiaris genomic,  
genomic survey sequence.  
ACCESSION CE153858  
VERSION CE153858.1 GI:35276514  
KEYWORDS GSS.  
SOURCE Canis familiaris (dog)  
ORGANISM Canis familiaris  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
1 (bases 1 to 685)  
Kirkness, E.F., Batina, V., Halpern, A.L., Levy, S., Remington, K.,  
Rusch, D.B., Delcher, A.L., Pop, M., Wang, W., Fraser, C.M. and  
Venter, J.C.  
The dog genome: survey sequencing and comparative analysis  
Science 301 (5641), 1898-1903 (2003)  
22875432  
MEDLINE  
PUBMED 14512627  
COMMENT Contact: Kirkness EF  
The Institute for Genomic Research  
Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,  
Rockville, MD 20850, USA  
Tel: 301-638-0200

Fax: 301-838-0208  
Email: [ekirknes@tigr.org](mailto:ekirknes@tigr.org)  
Class: shotgun.  
Location/Qualifiers  
1..685  
/organism="Canis familiaris"  
/mol\_type="genomic DNA"  
/strain="Standard Poodle"  
/db\_xref="taxon:9615"  
/clone\_lib="Dog Library"  
/note="Site 1: BcXI; Libraries were prepared from  
peripheral blood"

ORIGIN  
Query Match 79.1%; Score 17.4; DB 9; Length 685;  
Best Local Similarity 94.7%; Pred. No. 1e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GATCTTGACTGCTTTAA 19  
|||||  
DB 425 GATCTTGACTGCTTTAA 407

RESULT 12  
CG922134/c  
LOCUS CG922134 700 bp DNA linear GSS 12-DEC-2003  
DEFINITION MBETC39TF mch2 Medicago truncatula genomic clone 60H5, genomic  
survey sequence.  
ACCESSION CG922134  
VERSION CG922134.1 GI:39781812  
KEYWORDS GSS.  
SOURCE Medicago truncatula (barrel medic)  
ORGANISM Medicago truncatula  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;  
Medicago.  
1 (bases 1 to 700)  
Town, C.D., Shetty, J., Koo, H. and Feldblyum, T.F.  
Sequencing of BAC ends from Medicago truncatula  
Unpublished (2003)  
Other GSSs: MBETC39TR  
Contact: Chris Town  
TIGR  
9712 Medical Center Drive, Rockville, MD 20850, USA.  
Tel: 301-838-3523  
Fax: 301-838-0208  
Email: [cdtown@tigr.org](mailto:cdtown@tigr.org)  
Seq primer: TGTAAACGACGCGCCAGT  
Class: BAC ends.  
Location/Qualifiers  
1..700  
/organism="Medicago truncatula"  
/mol\_type="genomic DNA"  
/culivar="genotype AL7"  
/db\_xref="taxon:3880"  
/clone="60H5"  
/clone\_lib="mch2"  
/note="Vector: pBelobAC11; Site 1: HindIII; Site 2:  
HindIII; Cook, D.R. and Kim, D.U., unpublished"

ORIGIN  
Query Match 79.1%; Score 17.4; DB 9; Length 700;  
Best Local Similarity 94.7%; Pred. No. 1e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 ATCTTGACTGCTTTAA 20  
|||||  
DB 173 ATCTTGACTGCTTCAAC 155

RESULT 13  
AG452969

JOURNAL	single-nucleotide polymorphisms for mapping in Arabidopsis thaliana
MEDLINE	Genome Res. 13 (6), 1250-1257 (2003)
PUBMED	22683290
COMMENT	12799357
	Contact: Welschaar B
	ADIS DNA core facility at MPiz
	Max-Planck-Institute for Plant Breeding Research
	Carl-von-Linne Weg 10, 50829 Koeln, Germany
	Fax: 00492215062851
	Email: welschaar@mpiz-koeln.mpg.de
	Insert Length: 242 Std. Error: 0.00
	Plate: 15 row: 0 column: 01
	Seq primer: PB12; GGTGGCGGCGCCCTCTAG.
FEATURES	Location/Qualifiers
Source	1..242
	/organism="Arabidopsis thaliana"
	/mol_type="mRNA"
	/ecotype="C24"
	/db_xref="GAB1:601306"
	/db_xref="taxon:3702"
	/clone="MP12p76700115Q"
	/tissue_type="seedling"
	/dev_stage="few days old seedlings"
	/lab_host="E. coli XL1-Blue MFP"
	/clone_lib="MP12-ADIS-008"
	/note="Vector: Bluescript SK (-); Site 1: EcoRI, Site 2: XhoI; cDNA library from Arabidopsis thaliana, accession C24, seedling; lambda ZAPRI phage library was made at the Max-Planck-Institute of Molecular Plant Physiology, Golem, Germany and mass-excised at the Max-Planck-Institute for Plant Breeding Research, Cologne, Germany; cloning sites EcoRI-XhoI; Note: Sequencing granted in the context of the GABI Arabidopsis Verbund II: Genetic Diversity, 'Establishment of high-efficiency SNP-based mapping tools and development of methods for genome-wide mutation detection'; PI: Bernd Welschaar Sequence submission managed by RZPD/GABI-Primary database: <a href="http://gabi.rzpd.de">http://gabi.rzpd.de</a> This clone is available from RZPD, contact RZPD (clone@rzpd.de) for further information."
ORIGIN	
Query Match	78.2%; Score 17.2; DB 6; Length 242;
Best Local Similarity	86.4%; Pred. NO. 1.2e+03;
Matches	15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Cy	1 GATCTTGACTGCTCTTAACCC 22
Db	166 GATCTTGTCTTGTCTAACC 145
RESULT 15	
AU099200/c	300 bp mRNA linear EST 23-JAN-2004
LOCUS	AU099200 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone
DEFINITION	HRC01228 similar to Human homeobox-containing protein mRNA, mRNA
SEQUENCE	sequence.
ACCESSION	AU099200
VERSION	AU099200.1 GI:13550329
KEYWORDS	EST.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
	1 (bases 1 to 300)
REFERENCE	Suzuki,Y., Tainoda,T., Sese,J., Tajira,H., Mizushima-Sugano,J.,
AUTHORS	Hata,H., Ota,T., Isegai,T., Tanaka,T., Nakamura,Y., Suyama,A.,
	Sakaki,Y., Morishita,S., Okubo,K. and Sugano,S.
	Identification and characterization of the potential promoter
	regions of 1031 kinds of human genes
	Genome Res. 11 (5), 677-684 (2001)
TITLE	
JOURNAL	21235182
MEDLINE	21337467
PUBMED	
COMMENT	Contact: Yutaka Suzuki

## Department of Virology

Institute of Medical Science, University of Tokyo  
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan

Email: yusuzuki@ims.u-tokyo.ac.jp

Suzuki, Y., Yoshitomo-Nakagawa, K., Maruyama, K., Suyama, A. and  
Sugano, S. Construction and characterization of a full  
length-enriched and a 5'-end-enriched cDNA library. Gene 200 (1-2),  
149-156 (1997)

## FEATURES

## Source

Location/Qualifiers  
1..300  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="HRC01728"  
/clone\_lib="Sugano Homo sapiens cDNA library"

## ORIGIN

## Query Match

78.2%; Score 17.2; DB 1; Length 300;

Best Local Similarity 86.4%; Pred. No. 1.2e+03;

Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

## QY

1 GATCTGTACTGCTTAACCC 22

## DB

165 GATCTGTCTGCTTAACCTC 144

Search completed: February 8, 2005, 22:18:05  
Job time : 270.041 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 8, 2005, 11:28:00 / Search time 423.751 Seconds  
(without alignments)  
8481.439 Million cell updates/sec

Title: US-10-754-437-6

Perfect score: 76  
Sequence: 1 gattctgttcgtcgtccccc.....ccctgcataataatcaattc 76

Scoring table: IDENTITY\_NUC  
Gapop 10.0 / Gapext 1.0

Searched: 4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl: \*  
1: gb\_ba: \*  
2: gb\_hgt: \*  
3: gb\_in: \*  
4: gb\_om: \*  
5: gb\_ov: \*  
6: gb\_pac: \*  
7: gb\_ph: \*  
8: gb\_pl: \*  
9: gb\_pr: \*  
10: gb\_ro: \*  
11: gb\_srs: \*  
12: gb\_sy: \*  
13: gb\_un: \*  
14: gb\_vl: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	55.2	72.6	5562	2	AB091787	AB091787 Ocolemur.
2	54.2	71.3	212886	2	AC093946	AC093946 Rattus no
3	54.2	71.3	234471	2	AC121424	AC121424 Rattus no
4	52.6	69.2	363	11	BV089294	BV089294 RPAMMSR00
5	52.6	69.2	363	11	BV097602	BV097602 RPAMMSR00
6	52.6	69.2	463	10	DB306452	DB3064 Mus musculu
7	52.6	69.2	695	11	BV089295	BV089295 RPAMMSR00
8	52.6	69.2	695	11	BV097603	BV097603 RPAMMSR00
9	52.6	93.84	10	AF294397	AF294397 Mus muscu	
10	52.6	69.2	95826	10	AL805974	AL805974 Mouse DN
11	52	68.4	5684	9	AB091785	AB091785 Lemur cat
12	52	68.4	8004	9	AB091786	AB091786 Lemur cat
13	52	68.4	8810	9	AY040206	AY040206 Homo sapi
14	52	68.4	158142	2	AL357130	AL357130 Homo sapi
15	52	68.4	259202	2	AC002366	AC002366 Human Xp2
16	51	67.1	6465	9	AB091781	AB091781 Pan trogl
17	49.4	65.0	6451	4	AB091789	AB091789 Bos tauru
18	49	64.5	6442	4	AB091783	AB091783 Saimiri s
19	47.8	62.9	7425	4	AB091791	AB091791 Sus scrofu

20	47.2	62.1	5712	4	AB091793 Equus cab
21	45.6	60.0	7163	9	AB091782 Pan trogl
22	45.6	60.0	38765	9	BS000568 Pan trogl
23	45.6	60.0	177654	2	AP000918 Homo sapi
24	45.6	60.0	177726	9	AC145770 Pan trogl
25	45.6	60.0	190089	9	BS000576 Pan trogl
26	45.6	60.0	200214	9	AC013412 Homo sapi
27	44.2	58.2	5591	4	AB091784 Equus cab
28	42.4	55.8	5151	9	AB091788 Equus cab
29	41.8	55.0	296	10	S74899
30	41.8	55.0	727	10	MUSMAMEB
31	41.8	55.0	765	4	AB032194 Equus cab
32	41.8	55.0	789	4	AB032193 Equus cab
33	41.8	55.0	798	10	BC059090 Mus muscu
34	41.8	55.0	799	10	MUSMAMEA
35	41.8	55.0	6931	4	AB091792 Equus cab
36	41	53.9	6264	4	AB091790 Bos tauri
37	40.6	53.4	7454	9	AB091784 Saimiri s
38	40.2	52.9	65	6	COS31502 Sequence
39	40.2	52.9	399	10	RU07054 Rattus norv
40	40.2	52.9	457	10	RU060564 Rattus norv
41	40.2	52.9	476	6	AR452534 Sequence
42	40.2	52.9	753	10	RU060562 Rattus norv
43	40.2	52.9	780	10	U01245 Rattus norv
44	40.2	52.9	812	10	RU067130 Rattus norv
45	40.2	52.9	825	10	RU051195 Rattus norv

#### ALIGNMENTS

RESULT 1  
AB091787  
LOCUS  
Ocolemur garnettii AMELX gene for amelogenin, partial cds.  
ACCESSION  
AB091787  
VERSION  
AB091787.1 GI:29126027  
KEYWORDS  
SOURCE  
ORGANISM  
Ocolemur garnettii (small-eared galago)  
Ocolemur garnettii  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Strepsithini; Galagonidae; Ocolemur.

REFERENCE  
1  
Iwase, M., Satta, Y., Hirai, Y., Hirai, H., Imai, H. and Takahata, N.  
From the Cover: The amelogenin loci span an ancient pseudautosomal  
boundary in diverse mammalian species  
JOURNAL  
Proc. Natl. Acad. Sci. U.S.A. 100 (9), 5258-5263 (2003)  
MEDLINE  
22608569  
PubMed  
12672962

REFERENCE  
2 (bases 1 to 5562)  
Iwase, M., Satta, Y. and Takahata, N.  
Direct Submision  
Submitted (19-SEP-2002) Mineyo Iwase, Graduate University for  
Advanced Studies (Sokendai), Department of Biosystems Science;  
Shonan kokusaimura, Hayama, Kanagawa 240-0193, Japan  
(E-mail: iwase@minet.koryuwo1.soken.ac.jp, Tel: 81-468-58-1571,  
Fax: 81-468-58-1544)  
FEATURES  
source  
1..5562  
Location/Qualifiers  
/organism="Ocolemur garnettii"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:30611"  
/sex="male"  
267..5562  
/gene="AMELX"  
join(267..321,1401..1412)  
/gene="AMELX"  
join(1413..1466,3385..3432,4699..4740,4832..4876,  
5125..5562)  
/gene="AMELX"  
/codon\_start=1  
/product="amelogenin"  
/protein\_id="BAC66107.1"



Query Match 71.3%; Score 54.2; DB 2; Length 21286;  
 Best Local Similarity 94.9%; Pred. No. 1.1e-09;  
 Matches 56; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

1 GATTTGTTGGCTGCTGCTGAGACAGCCTTTGCTATGCCCGTAGTAATAATACCCC 59  
 112447 GATCTTGTTCCTGCTGCTGAGACAGCTTTTGTATGCCCGTAGTAATAATACCCC 112389

RESULT 3  
 AC121424 234471 bp DNA 1linear HTG 21-SEP-2002  
 LOCUS Rattus norvegicus clone CH230-122F17, \*\*\* SEQUENCING IN PROGRESS  
 DEFINITION \*\*\* 2 unordered pieces.  
 AC121424  
 AC121424.3 'GI:23265350  
 HTG; HTGS PHASE1; HTGS DRAFT; HTGS\_ENRICHED.  
 Rattus norvegicus (Norway rat)  
 SOURCE Rattus norvegicus  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.  
 1 (bases 1 to 234471)  
 Muzny,D,Marie, Metzker,M, Lee, A, Abramson, S., Adams, C., Alder, J.,  
 Allen, C., Allen, H., Albrooks, S., Amin, A., Anguiano, D.,  
 Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,  
 Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, W., Benahmed, F.,  
 Bissel, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,  
 Bryant, N., Bunay, C., Burch, P., Burrell, K., Calderon, E.,  
 Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,  
 Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,  
 Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,  
 Davila, M.L., Davis, C., Davy-Carroil, L., De Anda, C., Dederich, D.,  
 Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Ditya, K.,  
 Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,  
 Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,  
 Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,  
 Frazer, C.M., Gabisti, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,  
 Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, M., Guevara, W.,  
 Gundacker, P., Haaland, W., Hamill, C., Hamilton, K., Hamilton, K.,  
 Harvey, Y., Havlak, P., Haves, A., Henderson, N., Hernandez, J.,  
 Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogue, M.,  
 Hollins, B., Howells, S., Huylk, S., Hume, J., Idlebird, D., Jackson, A.,  
 Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,  
 Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,  
 Kowis, C., Kratt, C.L., Ledow, H., Levay, J., Lewis, L., Li, Z., Liu, J.,  
 Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,  
 Lorenshuber, L., Louised, H., Lozano, R.J., Lu, X., Ma, J.,  
 Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A.,  
 Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,  
 Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,  
 Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,  
 Morgan, M., Morris, K., Morris, S., Nandassa, M., Murphy, M., Nair, L.,  
 Nankervyls, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,  
 Nwackelmele, O., Okunolu, G., Olamunsgoon, A., Pal, S., Parks, K.,  
 Pascernek, S., Paul, H., Perez, A., Perez, L., Pfrankoch, C.,  
 Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L.,  
 Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,  
 Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,  
 Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,  
 Sanders, W., Savery, G., Scherer, S., Scott, G., Shltsman, S., Shen, H.,  
 Shetty, J., Shvartsbeyn, A., Sison, I., Sitter, C.D., Smajls, D.,  
 Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Soza, J.,  
 Steimle, M., Strong, R., Sutton, A., Svatek, A., Taboi, P., Taylor, C.,  
 Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umanu, K.,  
 Vales, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J.,  
 Wang, O., Wang, S., Warren, J., Warren, R., Wei, X., White, P.,  
 Williams, G., Willson, R., Wleczek, R., Wooden, H., Worley, K.,  
 Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,  
 Yu, F., Zhang, V., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von  
 Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,  
 Weinstock, G., and Gibbs, R.A.  
 Direct Submission

JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

Unpublished  
 2 (bases 1 to 234471)  
 Worley, K.C.  
 Direct Submission  
 Submitted (18-MAY-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 234471)  
 Rat Genome Sequencing Consortium.  
 Direct Submission  
 Submitted (21-SEP-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

On Sep 21, 2002 this sequence version replaced gi:21908624.  
 The sequence in this assembly is a combination of BAC based reads  
 and whole genome shotgun sequencing reads assembled using Atlas  
 (http://www.hgsc.bcm.tmc.edu/projects/rat/). As a result, the  
 sequence may extend beyond the end of the clone and there may be  
 contigs that consist entirely of whole genome shotgun sequence  
 reads. Both end sequences and whole genome shotgun sequence only  
 contigs will be indicated in the feature table.

----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: http://www.hgsc.bcm.tmc.edu/  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: GX01  
 Center clone name: CH230-122F17  
 ----- Summary Statistics  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 220465 bases at least Q40  
 Consensus quality: 222968 bases at least Q30  
 Consensus quality: 224416 bases at least Q20  
 Estimated insert size: 241282; sum-of-contigs estimation  
 Quality coverage: 4x in Q20 bases; sum-of-contigs estimation

-----  
 \* NOTE: Estimated insert size may differ from sequence length  
 \* (see http://www.hgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 2 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 232989: contig of 232989 bp in length  
 \* 232990 233089: gap of unknown length  
 \* 233090 234471: contig of 1382 bp in length.  
 Location/Qualifiers  
 1. 234471  
 /organism="Rattus norvegicus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10116"  
 /clone="CH230-122F17"  
 104912..105798  
 /note="clone boundary  
 clone end: T7  
 site: EcoRI  
 end\_sequence: BH268493"  
 complement(231872..232742)  
 /note="clone boundary  
 clone end: Sp6  
 site: EcoRI  
 end\_sequence: BH268494"

ORIGIN  
 Query Match 71.3%; Score 54.2; DB 2; Length 234471;  
 Best Local Similarity 94.9%; Pred. No. 1.1e-09;  
 Matches 56; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

1 GATTTGTTGGCTGCTGCTGAGACAGCCTTTGCTATGCCCGTAGTAATAATACCCC 59







Promoter  
TATA\_signal  
mRNA  
CDS

/chromosome="X"  
1..8546  
7188..7194  
8547..8612  
/product="amelogenin"  
8559..8612  
/note="enamel matrix protein"  
/codon\_start=1  
/product="amelogenin"  
/protein\_id="AA010078.1"  
/db\_xref="GI:965406"  
/translation="MGTWILFACLLGAPAMP"

## ORIGIN

Query Match 69.2%; Score 52.6; DB 10; Length 9384;  
Best Local Similarity 93.2%; Pred. No. 2.7e-09;  
Matches 55; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GATTTTGTTCCTGCTCTGAGAGACGCTTTGCTATGCCCGTAGTAATACCCC 59  
|||||  
Db 8570 GATTTTGTTCCTGCTCTGAGAGACGCTTTGCTATGCCCGTAGTAATACCCC 8628

RESULT 10  
AL805974 95826 bp DNA linear ROD 29-AUG-2003  
LOCUS AL805974/c  
DEFINITION Mouse DNA sequence from clone RP23-334F21 on chromosome X, complete sequence.  
ACCESSION AL805974  
VERSION AL805974.8 GI:34366495  
KEYWORDS HTG.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 95826)  
Chapman, J.  
Direct Submission  
Submitted (28-AUG-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
On Aug 30, 2003 this sequence version replaced gi:25955748.  
Sequence from the Mouse Genome Sequencing Consortium whole genome shotgun may have been used to confirm this sequence. Sequence data from the whole genome shotgun alone has only been used where it has a phred quality of at least 30.

-----Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: <http://www.sanger.ac.uk>  
Contact: [humquery@sanger.ac.uk](mailto:humquery@sanger.ac.uk)  
-----

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.  
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.  
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) RP23-334F21 is from the RPI-23 Mouse BAC library constructed by the group of Pieter de Jong.

For further details see <http://www.chori.org/bacpac/home.htm>  
VECTOR: pBAC3.6  
Location/Qualifiers  
1..95826  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosome="X"  
/clone="RP23-334F21"  
/clone\_11b="RPI-23"

## ORIGIN

Query Match 69.2%; Score 52.6; DB 10; Length 95826;  
Best Local Similarity 93.2%; Pred. No. 4.2e-09;  
Matches 55; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GATTTTGTTCCTGCTCTGAGAGACGCTTTGCTATGCCCGTAGTAATACCCC 59  
|||||  
Db 22692 GATTTTGTTCCTGCTCTGAGAGACGCTTTGCTATGCCCGTAGTAATACCCC 22634

RESULT 11  
AB091785 5684 bp DNA linear PRI 02-MAY-2003  
LOCUS AB091785  
DEFINITION Lemur catca AMELX gene for amelogenin, partial cds.  
ACCESSION AB091785  
VERSION AB091785.1 GI:29126023  
KEYWORDS  
SOURCE Lemur catca (ring-tailed lemur)  
ORGANISM Lemur catca  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Strepsirrhini; Lemnidae; Lemur.  
1  
Iwase, M., Satta, Y., Hirai, Y., Hirai, H., Imai, H. and Takahata, N.  
From the Cover: The amelogenin loci span an ancient pseudautosomal boundary in diverse mammalian species  
Proc. Natl. Acad. Sci. U.S.A. 100 (9), 5258-5263 (2003)  
22608569  
12672962  
2 (bases 1 to 5684)  
Iwase, M., Satta, Y. and Takahata, N.  
Direct Submission  
Submitted (19-SEP-2002) Mineyo Iwase, Graduate University for Advanced Studies (Sokendai), Department of Biosystems Science; Shonan Kokusai-mura, Hayama, Kanagawa 240-0193, Japan  
(E-mail: iwase@nsl.soken.ac.jp, Tel: 81-468-58-1571, Fax: 81-468-58-1544)  
FAX: 81-468-58-1544

FEATURES  
source  
Location/Qualifiers  
1..5684  
/organism="Lemur catca"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9447"  
/sex="male"  
100..5684  
/gene="AMELX"  
join(100..156,1503..1514)  
/gene="AMELX"  
join(1515..1568,3465..3512,4804..4845,4937..4981,5229..55684)  
/gene="AMELX"  
/codon\_start=1  
/product="amelogenin"  
/protein\_id="BAC66105.1"  
/db\_xref="GI:29126023"  
/translation="MGTWILFACLLGAPAMP"

gene  
5'UTR  
CDS  
100..5684  
/gene="AMELX"  
join(100..156,1503..1514)  
/gene="AMELX"  
join(1515..1568,3465..3512,4804..4845,4937..4981,5229..55684)  
/gene="AMELX"  
/codon\_start=1  
/product="amelogenin"  
/protein\_id="BAC66105.1"  
/db\_xref="GI:29126023"  
/translation="MGTWILFACLLGAPAMP"

## ORIGIN

Query Match 68.4%; Score 52; DB 9; Length 5684;  
Best Local Similarity 91.7%; Pred. No. 4.2e-09;  
Matches 55; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 GATTTGTTGCTGCTGCTGAGACGCTTGTCTATGCCGTGAGTAAATACCCCT 60  
 DB 1526 GATTTGTTGCTGCTGCTGAGACGCTTGTGCGCATGCGCGTGAATAACACCCCT 1585

RESULT 12  
 AB091786 8004 bp DNA linear PRI 02-MAY-2003  
 LOCUS Lemur catla AMELY gene for amelogenin, partial cds.  
 DEFINITION AB091786  
 ACCESSION AB091786.1 GI:29126025  
 VERSION  
 KEYWORDS  
 SOURCE Lemur catla (ring-tailed lemur)  
 ORGANISM  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Strepsitrhini; Lemnridae; Lemur.  
 REFERENCE 1  
 Iwase, M., Saita, Y., Hirai, Y., Hirai, H., Imai, H. and Takahata, N.  
 From the Cover: The amelogenin loci span an ancient pseudautosomal  
 boundary in diverse mammalian species  
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 100 (9), 5258-5263 (2003)  
 MEDLINE 12672962  
 PUBMED 12672962  
 REFERENCES 2 (bases 1 to 8004)  
 Iwase, M., Saita, Y. and Takahata, N.  
 Direct Submision  
 Submitted (19-SEP-2002) Mineyo Iwase, Graduate University for  
 Advanced Studies (Sokenai), Department of Biosystems Science;  
 Shonan Kokusai-mura, Hayama, Kanagawa 240-0193, Japan  
 (E-mail: iwase@koryu.w01.soken.ac.jp, Tel: 81-468-58-1571,  
 Fax: 81-468-58-1544)  
 FEATURES  
 source  
 Location/Qualifiers  
 1..8004  
 /organism="Lemur catla"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9447"  
 /sex="male"  
 1382..6505  
 /gene="AMELY"  
 join(1382..1437,3040..3051)  
 /gene="AMELY"  
 join(3052..3105,5192..5239,6461..6505)  
 /gene="AMELY"  
 /codon\_start=1  
 /product="amelogenin"  
 /protein\_id="BA066106.1"  
 /db\_xref="GI:29126026"  
 /translation="MGTWILFACLLGAAPAMPPLQPHRGHYINFSYVLTPLKMYOS  
 MTRPP"  
 ORIGIN  
 Query Match 68.4%; Score 52; DB 9; Length 8004;  
 Best Local Similarity 91.7%; Pred. No. 4.5e-09;  
 Matches 55; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 GATTTGTTGCTGCTGCTGAGACGCTTGTCTATGCCGTGAGTAAATACCCCT 60  
 DB 3063 GATTTGTTGCTGCTGCTGAGACGCTTGTGCGCATGCGCGTGAATAACACCCCT 3122

RESULT 13  
 AY040206 8810 bp DNA linear PRI 10-APR-2002  
 LOCUS Homo sapiens amelogenin precursor (AMELX) gene, complete cds.  
 DEFINITION AY040206  
 ACCESSION AY040206.1 GI:15028582  
 VERSION  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 8810)

AUTHORS Hart, P.S., Hart, T.C., Simmer, J.P. and Wright, J.T.  
 TITLE A nomenclature for X-linked amelogenesis imperfecta  
 JOURNAL Arch. Oral Biol. 47 (4), 255-260 (2002)  
 MEDLINE 21920287  
 PUBMED 11922868  
 REFERENCES 2 (bases 1 to 8810)  
 Hart, S., Hart, T.C. and Wright, J.T.  
 Direct Submision  
 Submitted (14-JUN-2001) Human Genetics, University of Pittsburgh,  
 3550 Terrace St., 572A Scaife Hall, Pittsburgh, PA 15090, USA  
 Location/Qualifiers  
 1..8810  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="X"  
 /map="Xp22.3-p22.1"  
 1..8791  
 /gene="AMELX"  
 1..1416  
 /gene="AMELX"  
 1386..1391  
 /gene="AMELX"  
 join(1417..1472,2781..2846,4783..4830,6240..6281,  
 6552..6977,8611..8791)  
 /gene="AMELX"  
 /product="amelogenin precursor"  
 1417..1472  
 /gene="AMELX"  
 /number=1  
 2781..2846  
 /gene="AMELX"  
 /number=2  
 join(2793..2846,4783..4830,6240..6281,6552..6977,  
 8611..8616)  
 /gene="AMELX"  
 /codon\_start=1  
 /product="amelogenin precursor"  
 /protein\_id="AAK77213.1"  
 /db\_xref="GI:15028583"  
 /translation="MGTWILFACLLGAAPAMPPLQPHRGHYINFSYVLTPLKMYOS  
 TRPPSYGVEBPMGWLHNOIIVLSOGRPTHTLOHHIIPVRAOQVITPOQPMRP  
 VPGQSMPTIOHNOIRLPRRAOQVPOQVQPHQPMQROPRVNHMOVLPRPRLRP  
 MPQPLRPMPLDLTLEAMPSTDKTRRESVD"  
 2793..2840  
 /gene="AMELX"  
 4783..4830  
 /gene="AMELX"  
 /number=3  
 6107..6148  
 /gene="AMELX"  
 /number=4  
 6240..6281  
 /gene="AMELX"  
 /number=5  
 6552..6977  
 /gene="AMELX"  
 /number=6  
 8611..8791  
 /gene="AMELX"  
 /number=7  
 8785..8791  
 /gene="AMELX"  
 ORIGIN  
 Query Match 68.4%; Score 52; DB 9; Length 8810;  
 Best Local Similarity 91.7%; Pred. No. 4.5e-09;  
 Matches 55; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 GATTTGTTGCTGCTGCTGAGACGCTTGTCTATGCCGTGAGTAAATACCCCT 60  
 DB 2804 GATTTATTTGCTGCTGCTGAGACGCTTGTGCGCATGCGCGTGAATAACACCCCT 2863

RESULT 14  
AL357130/ 158142 bp DNA linear HTG 10-JUL-2001  
LOCUS Homo sapiens chromosome X clone RP13-169E15, 4 unordered pieces.  
DEFINITION AL357130.3 GI:9863814  
ACCESSION HTG; HTGS PHASE1; HTGS\_CANCELED.  
VERSION HTG; HTGS PHASE1; HTGS\_CANCELED.  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1  
McLay, K.  
Direct Submission  
Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk  
requester: clonerequest@sanger.ac.uk  
On Aug 21, 2000 this sequence version replaced gi:9214076.  
----- Genome Center  
Center: Sanger Centre  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: humquerry@sanger.ac.uk  
----- Project Information  
Center project name: DB169E15  
----- Summary Statistics  
Assembly program: XGAP4; version 4.5  
Sequencing vector: plasmid, L08752; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Consensus quality: 155448 bases at least Q40  
Consensus quality: 156466 bases at least Q30  
Consensus quality: 157038 bases at least Q20  
Insert size: 157842; sum-of-contigs  
Insert size: 160705; 33.4% error; agarose-fp  
Quality coverage: 4.56x in Q20 bases; sum-of-contigs Quality  
coverage: 4.56x in Q20 bases; agarose-fp  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 4 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\*  
\* 1 68897: contig of 68897 bp in length  
\* 68898 68897: gap of 100 bp  
\* 68898 122842: contig of 53845 bp in length  
\* 122843 122942: gap of 100 bp  
\* 122943 129584: contig of 6642 bp in length  
\* 129585 129684: gap of 100 bp  
\* 129685 158142: contig of 28458 bp in length.  
Location/Qualifiers  
1. 158142  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="X"  
/clone="RP13-169E15"  
/clone\_1fb="RP13-13.1"  
1. 68897  
/note="assembly\_fragment:01108  
fragment\_chain:1  
clone\_end:17  
vector\_side:left"  
68898. 122842  
/note="assembly\_fragment:00101  
fragment\_chain:1"  
122943. 129584  
/note="assembly\_fragment:00361  
fragment\_chain:1"  
129685. 158142  
misc\_feature  
misc\_feature  
misc\_feature

ORIGIN  
Query Match 68.4%; Score 52; DB 2; Length 158142;  
Best Local Similarity 91.7%; Pred. No. 7.9e-09;  
Matches 55; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
Y 1 GATTTTGTTGGCTGCTGCTGAGACGCTTGTCTATGCGGTGAGTAATACCCCT 60  
DB 44314 GATTTTATTTGGCTGCTGCTGAGACGCTTGTCTATGCGGTGAGTAATACCCCT 44255  
/note="assembly\_fragment:01285  
fragment\_chain:1  
clone\_end:8P6  
vector\_side:right"  
RESULT 15  
AC002366 259202 bp DNA linear PRI 11-JUN-1998  
LOCUS Human Xp22 BAC CP-285115 (from Caltech/Research Genetics), PAC  
DEFINITION RP11-27C22 (from Roswell Park Cancer Center), and Cosmid U3B5  
(from Lawrence Livermore), complete sequence.  
ACCESSION AC002366 U79549 U70036  
VERSION AC002366.1 GI:2739349  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 259202)  
Muzny, D., Ansari-Lari, M.A., Timme, K.M., Yu, W., Dugan, S., Lu, J.,  
Shen, Y., Rowland, K., Liu, W., Perez, L., Ding, Y., Haywood, M.,  
Jain, A., Leal, B., Logan, O., Nguyen, V., Savage, L., Shen, H.,  
Worley, K., Chen, F., Forcum, J., Atkinson, A.D., Chiu, M.W.,  
Correia, U.H., Brundage, S., Di, W., Chnault, C., Nelson, D. and  
Gibbs, R.A.  
Direct Submission  
Unpublished  
2 (bases 1 to 259202)  
Chiu, M.W.  
Direct Submission  
Submitted (23-JUL-1997) Molecular and Human Genetics, Baylor  
College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 259202)  
Chiu, M.W.  
Direct Submission  
Submitted (02-JUN-1998) Molecular and Human Genetics, Baylor  
College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
4 (bases 1 to 259202)  
Chiu, M.W.  
Direct Submission  
Submitted (11-JUN-1998) Molecular and Human Genetics, Baylor  
College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Jan 2, 1998 this sequence version replaced gi:2642176.  
Sequencing is completed to a minimum standard of double strand  
coverage with a minimum of 2 clones and 2 reads with no ambiguities  
or 2 chemistries with a minimum of 2 clones and 3 reads with no  
ambiguities. If the sequence quality does not meet this standard,  
it will be indicated in the annotation.  
The repeat regions shown were identified using RepeatMasker by  
Adrian Smit.  
Sequence similarities were identified using Powerblast by Jinghui  
Zhang.  
Exon/Intron boundaries of identified genes were chosen if there  
were canonical splice junctions that maintained sequence continuity  
across the splice junctions.  
Location/Qualifiers  
1. 259202  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
FEATURES  
source

misc_feature	/chromosome="X" /map="Xp22" /clone="Ct-285115, RPTC11-27C22, U35B5" complement(1..147) /note="overlap 1..147 in AC004467"
repeat_region	35..127 /rpt_family="MIR" complement(1087..1502) /rpt_family="L1" complement(5292..5579) /rpt_family="Alu" 5302..5553
repeat_region	/rpt_family="SVA" complement(6009..6063) /rpt_family="L1MC2" complement(7092..7282) /rpt_family="L1MA5" 7423..7721
repeat_region	/rpt_family="Alu" 9771..10072 /rpt_family="L1MC2" complement(10072..11146) /rpt_family="THR" complement(11156..11528) /rpt_family="THE1b" complement(12944..13151) /rpt_family="MER20" 13862..14010 /rpt_family="MLT1f" 14063..14352
repeat_region	/rpt_family="MER42a" 15847..16170 /rpt_family="Alu" complement(18743..19117) /rpt_family="MER7" 23507..23614 /rpt_family="L1" 23615..24425
repeat_region	/rpt_family="L1PA11" complement(25475..25968) /rpt_family="MLT1d" 27234..27683 /note="similar to BST with GenBank Accession Number H48827"
repeat_region	/db_xref="dbSTS:G28372" complement(28186..28366) /rpt_family="MER20" 28715..28995 /rpt_family="Alu" complement(30217..30408) /rpt_family="L1MA5" 31199..31364 /rpt_family="Alu" complement(33939..34460) /rpt_family="L1MA2" 34952..35384 /rpt_family="MLT1d" 36035..36126 /rpt_family="MIR2" 36850..37035
repeat_region	/rpt_family="MER20" 38042..38250 /rpt_family="MER20" 38132..38510 /rpt_family="MSTa" 38797..40103 /rpt_family="MIR" complement(42465..42539) /rpt_family="MIR" 47139..47335 /rpt_family="MER3" complement(48993..49166) /rpt_family="MIR"

```

repeat_region complement (49718. .44918)
repeat_region /rpt_family="L1PA11"
repeat_region complement (51943. .52505)
repeat_region /rpt_family="L1MB3"
repeat_region 53226. .53677
repeat_region /rpt_family="MLT1d"
repeat_region complement (54846. .54966)
repeat_region /rpt_family="MIR"
repeat_region 55040. .55226
repeat_region /rpt_family="MER20"
repeat_region 55691. .55837
repeat_region /rpt_family="Alu"
repeat_region 56051. .56486
repeat_region /rpt_family="L1MB3a"
repeat_region 58557. .59030
repeat_region /rpt_family="MLT1d"
repeat_region complement (59270. .59428)
repeat_region /rpt_family="MER5"
repeat_region complement (59847. .60030)
repeat_region /rpt_family="MER20"
repeat_region complement (61290. .61384)
repeat_region /rpt_family="L1PB1"
repeat_region 61683. .61811
repeat_region /rpt_family="MER20"
repeat_region 61907. .62113
repeat_region /rpt_family="L1MB7"
repeat_region complement (63588. .63969)
repeat_region /rpt_family="L1MA10"
repeat_region 64800. .65073
repeat_region /rpt_family="Alu"
repeat_region complement (67157. .67209)
repeat_region /rpt_family="MIR"
repeat_region 67637. .75845
repeat_region /gene="HMBELX"
repeat_region 67637. .68496
repeat_region /gene="HMBELX"
repeat_region /note="similar to bovine ameloogenin promoter"
repeat_region join (68495. .68552,69861. .69926,71863. .71910,73320. .73361,73613. .74057,75691. .75845)
repeat_region /gene="HMBELX"
repeat_region /product="amelogenin"
repeat_region join (69873. .69926,71863. .71910,73320. .73361,73613. .74057,75691. .75845)
repeat_region /gene="HMBELX"
repeat_region /note="similar to amelogenin encoded by GenBank Accession Number M86932"
repeat_region /codon_start=1
repeat_region /product="amelogenin"
repeat_region /protein_id="AAC21581.1"
repeat_region /db_xref="GI:2739350"
repeat_region /translation="KGTWILFACILGAFAFMRPLRPNGRGYINFSYEVLRLTKWOS
VBRPYSVGVEPMGGLIHNQIIVLSQNRPTNLQVNNIIVYDAQQRVLRQQRMR
VYQGSKMTPIQHNORPLRPRAQQRVQRFQVROHNRQQRVVRNMRPLRPPRLRP
MRWQPLRPLRDLTLKMRPSTDKTKREVD"
repeat_region complement (77161. .77331)
repeat_region /rpt_family="MER3"
repeat_region complement (79184. .79473)
repeat_region /rpt_family="Alu"
repeat_region complement (79881. .80169)
repeat_region /rpt_family="Alu"
repeat_region 80196. .80253
repeat_region /rpt_family="MER20"
repeat_region 80384. .80459
repeat_region /rpt_family="L1PB3"
repeat_region complement (81544. .81633)
repeat_region /rpt_family="MIR"
repeat_region 84789. .84957
repeat_region /rpt_family="L1MA2"
repeat_region 84976. .85252
repeat_region /rpt_family="Alu"
repeat_region 85294. .85692

```

Wed Feb 9 09:28:18 2005

us-10-754-437-6.rge

Page 10

	Beat	Local	Similarity	91.7%	Pred:	No. 8.7e-09			
	Matches	55	Conservative	0	Mismatches	5	Indels	0	Gaps
Qy	1	GATTTGTTGGCCCTGCCTCCTGGAGACAGCCTTGCATATCCCGTAGTAAATACCCCT	60						
Db	69884	GATTTATTTGGCCCTGCCTCCTGGAGACAGCCTTGGCATTCGTGATTAACACACCCCT	69943						

Search completed: February 8, 2005, 16:36:54  
Job time : 426.751 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 8, 2005, 11:25:10 ; Search time 100.817 Seconds  
(without alignment)  
3957.220 Million cell updates/sec

Title: US-10-754-437-6

Perfect score: 76  
Sequence: 1 gatctgttcgtcgtccccc.....ccctgcataatcattcattc 76

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4134886 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_23Sep04:\*  
1: geneseqn1980s:\*  
2: geneseqn1990s:\*  
3: geneseqn2000s:\*  
4: geneseqn2001s:\*  
5: geneseqn2001bs:\*  
6: geneseqn2002as:\*  
7: geneseqn2002bs:\*  
8: geneseqn2003as:\*  
9: geneseqn2003bs:\*  
10: geneseqn2003cs:\*  
11: geneseqn2003ds:\*  
12: geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	52	68.4	259202	12	ADQ18492 Human sof
2	40.4	53.2	270	12	ADM80835 Human sof
3	40.4	53.2	549	12	ADM80826 Human CAD
4	40.4	53.2	623	12	ADM80827 Human CAD
5	40.2	52.9	65	6	ABN28389 Rat splic
6	40.2	52.9	476	3	AAZ50832 Rat amelo
7	39.2	51.6	556	10	ADB59026 Toxicity-
8	39.2	51.6	556	10	ADB53782 Primary r
9	38.8	51.1	722	6	AA141111 GAML rela
10	38.8	51.1	752	6	AA141110 GAML rela
11	37.2	48.9	318	12	ADM80836 Human CAD
12	37.2	48.9	750	2	AAZ07020 Amelogeni
13	37.2	48.9	793	2	AAZ07018 Amelogeni
14	35.6	46.8	802	2	AAZ07019 Amelogeni
15	35.6	46.8	852	12	ADQ22977 Human sof
16	27.6	36.3	1183	6	AB158186 Murine Sc
17	27.6	36.3	1181	6	AB158182 Murine Sc
18	26.8	35.3	849	12	ADJ48146 Maize c11
19	26.6	35.0	4031	4	AB103527 Drosophila
20	26.6	35.0	4940	4	AB103510 Drosophila
21	26.6	35.0	8256	4	AB103526 Drosophila

22	26.6	35.0	198161	6	ABK83564 Human CDN
23	26.6	35.0	198161	12	ADQ17348 Human sof
24	26	34.2	795	6	AB190186 Human pol
25	26	34.2	930	10	ADK68445 Simian T-
26	26	34.2	1802	2	AAT05712 STLV-PH96
27	26	34.2	3850	2	AAT05713 PH96-PTL
28	25.8	33.9	4713	8	ABX34656 Human mdd
29	25.8	33.9	5430	5	AA585365 DNA encod
30	25.6	33.7	936	11	ABD00916 Klebsiell
31	25.4	33.4	591	2	AA212126 Neisseria
32	25.4	33.4	591	3	AA254407 Neisseria
33	25.4	33.4	1137	2	AA253817 Neisseria
34	25.4	33.4	2000	8	ADA73001 Rice gene
35	25.4	33.4	77287	9	AA582821 Murine tu
36	25.4	33.4	167739	9	AA582821 Murine tu
37	25.2	33.2	274	10	ABX84738 Corn ear-
38	25.2	33.2	327	10	ABX87256 Corn ear-
39	24.6	32.4	350	4	AA11877 Human bre
40	24.6	32.4	351	4	AA120767 Human bre
41	24.6	32.4	7466	4	AB109192 Drosophila
42	24.6	32.4	10825	6	ABN80207 Human che
43	24.6	32.4	32172	4	AA103473 Human rep
44	24.4	32.1	313	4	AA119256 Human bre
45	24.4	32.1	1410	10	ADC77539 Human HKE

## ALIGNMENTS

RESULT 1	ADQ18492	standard; DNA; 259202 BP.
ID	ADQ18492	
XX	ADQ18492;	
XX	26-AUG-2004 (first entry)	
XX	Human soft tissue sarcoma-upregulated DNA - SEQ ID 1311.	
XX	soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;	
XX	ds.	
XX	Homo sapiens.	
OS	26-NOV-2003; 2003WO-US038193.	
XX	26-NOV-2003; 2003WO-US038193.	
XX	26-NOV-2002; 2002US-0429739P.	
XX	(PROT-) PROTEIN DESIGN LABS INC.	
XX	Aziz N, Ginsburg WM, Zlotnick A;	
XX	WPI; 2004-441208/41.	
XX	Early detection of soft tissue sarcoma comprises determining expression	
XX	of a gene in a first soft tissue sample and a normal soft tissue sample	
XX	and comparing the gene expression, also useful in treating soft tissue	
XX	sarcoma.	
XX	Example 2; SEQ ID NO 1311; 210pp; English.	
XX	The invention relates to a novel method for detecting soft tissue sarcoma	
XX	which comprises obtaining a first soft tissue sample from an individual	
XX	and a normal soft tissue sample from the same or different individual,	
XX	determining the expression of a gene in both samples and comparing the	
XX	expression of the gene in both soft tissue samples, where a higher level	
XX	of protein expression in the first soft tissue sample indicates the	
XX	presence of soft tissue sarcoma. The method of the invention has	
XX	cytostatic applications and may be useful for detecting soft tissue	





CC	neurological, developmental, connective tissue and cell proliferative disorders including cancer, e.g. breast, prostate, ovarian, lung or colon cancer, obesity and Tangier disease.
XX	
SQ	Sequence 549 BP; 134 A; 203 C; 112 G; 100 T; 0 U; 0 Other;
Query Match	53.2%; Score 40.4; DB 12; Length 549;
Best Local Similarity	88.0%; Pred. No. 3.3e-05;
Matches 44;	Conservative 0; Mismatches 6; Indels 0; Gaps 0
OY	1 GATTTCGTTGGCTGCCTCCTGGAGACAGCTTGTATGCCCGTAGTA 50       47 GATTTCATTTGCTGCTGCTCTCGGAGCAGCTTTTGCAATCCATGCTTA 96
Db	
RESULT 4	
ID	ADM80827 standard; cDNA; 623 BP.
XX	
AC	ADM80827;
XX	
DT	03-JUN-2004 (first entry)
XX	
DE	Human CADECm-14 encoding cDNA SEQ ID NO:56.
KW	human; cell adhesion and extracellular matrix protein; CADECm;
KW	neuroprotective; cytosolic; anorectic; immune disorder;
KW	neurological disorder; developmental disorder;
KW	connective tissue disorder; cell proliferative disorder; cancer; obesity;
KW	Tangier disease; gene; ss.
XX	
OS	Homo sapiens.
XX	
FH	Key Location/Qualifiers
FT	CDS 36..563
FT	/tag= "A"
FT	/product= "CADECM-14"
XX	
PV	WO2004015396-A2.
XX	
PD	19-FEB-2004.
XX	
PF	12-AUG-2003; 2003MO-US025418.
XX	
PR	13-AUG-2002; 2002US-0403781P.
PR	30-AUG-2002; 2002US-0407034P.
PR	13-SEP-2002; 2002US-0410566P.
PR	24-SEP-2002; 2002US-0413482P.
PR	25-SEP-2002; 2002US-0413890P.
PR	08-NOV-2002; 2002US-0424904P.
PR	13-NOV-2002; 2002US-0426222P.
XX	
PA	(INCY-) INCYTE CORP.
XX	
P1	Elliott VS, Khare R, Emerling BM, Kable AE, Tran UK, Jin P;
P1	Beeha SD, Margulis JP, Swarnakar A, Chawla NK, Ramkumar J;
P1	Hafellia AA, Lee SY, Jiang X, Jackson AA, Richardson TW, Blake JJ;
P1	Mang JT, Chlen D, Yang YC;
DR	WPI; 2004-191795/18.
DR	P-PDB; ADM80785.
XX	
PT	New cell adhesion and extracellular matrix proteins, useful in
PT	diagnosing, treating and preventing immune, neurological, developmental,
PT	connective tissue and cell proliferative disorders including cancer.
XX	
PS	Claim 5; SEQ ID NO 56; 272pp; English.
XX	
CC	The present sequence encodes a human cell adhesion and extracellular
CC	matrix protein designated CADECm. CADECm sequences has neuroprotective,
CC	cytostatic and anorectic activities. The CADECm polypeptides and
CC	polynucleotides are useful in diagnosing, treating and preventing immune,
CC	neurological, developmental, connective tissue and cell proliferative

[illegible]

CC directly from WIFO at ftp.wifo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 65 BP; 8 A; 23 C; 14 G; 20 T; 0 U; 0 Other;  
Query Match 52.9%; Score 40.2; DB 6; Length 65;  
Best Local Similarity 93.3%; Pred. No. 1.9e-05;  
Matches 42; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 GATTTCCTTGGCCCTCTGGAGAGACCTTGTATGCCCT 45  
DB 6 GATCTGTTCCTGCTCTGGAGAGACCTTGTATGCCCT 50  
RESULT 6  
AAZ50832  
ID AAZ50832 standard; DNA, 476 BP.  
XX  
AC AAZ50832;  
XX  
DT 31-MAY-2000 (first entry)  
XX  
DE Rat amelogenin gene (A4).  
XX  
KW Amelogenin; splice variant; rat; (A4); chondrogenesis; osteogenesis;  
KW chondrogenic inducing molecule; CIM; cartilage growth; osteopathic;  
KW extracellular matrix protein; tooth enamel; enamel mineralisation;  
KW ameloblast; bone regeneration; composite cell construct; da.  
XX  
OS Rattus sp.  
XX  
FH Key Location/Qualifiers  
FT exon 1..36  
FT /\*tag= a  
FT /\*partial  
FT /\*number= 1  
FT 37..101  
FT /\*tag= b  
FT /\*number= 2  
FT 48..317  
FT /\*tag= c  
FT /\*product= "Rat amelogenin protein"  
FT /\*transl\_except= (pos:69..71, aa:Gly)  
FT 48..95  
FT /\*tag= d  
FT /\*tag= e  
FT /\*tag= e  
FT /\*label= Mature\_rat\_amelogenin\_protein  
FT 102..149  
FT /\*tag= f  
FT /\*number= 3  
FT 150..191  
FT /\*tag= g  
FT /\*number= 4  
FT 192..236  
FT /\*tag= h  
FT /\*number= 5  
FT 237..311  
FT /\*tag= i  
FT /\*number= 6d  
FT /\*note= "Comprises of gene segments 6a, b, c and d"  
FT 312..317  
FT /\*tag= j  
FT /\*number= 7  
FT /\*note= "Includes the stop codon"  
XX  
PN WO20006734-A1.  
XX  
PD 10-FEB-2000.  
XX  
PF 29-JUL-1999; 99WO-US017342.  
XX  
PR 29-JUL-1999; 98US-0094489P.  
XX

PA (NOUN ) UNTV NORTHWESTERN.  
XX  
PI Vels A, Nebgen DR;  
XX  
DR WPI; 2000-205464/18.  
XX  
DR P-PDSB; AAY45074.  
XX  
PT Novel amelogenin polypeptides and polynucleotides, useful for enhancing  
PT bone generation in mammals and synthesizing bone matrix or articular  
PT surfaces at implant sites.  
XX  
PS Example 2; Fig 11B; 79pp; English.  
XX  
CC The present DNA sequence is the full-length rat amelogenin gene (A4),  
CC comprising exons 1-7, including the exon segment 6d. It is derived from  
CC the rat incisor odontoblast-pulp cDNA library. The splice variants of  
CC this gene functions as an osteogenic or chondrogenic inducing molecule  
CC (CIM), which is useful for enhancing bone or cartilage growth. It has  
CC osteopathic activity. Amelogenin belongs to the family of extracellular  
CC matrix proteins, in developing tooth enamel, that are produced by the  
CC ameloblasts and plays a role in enamel mineralisation. Chondrogenic or  
CC osteogenic inducing amelogenin molecules are useful to induce  
CC differentiation of cells to the osteogenic and chondrogenic phenotypes  
CC and can be used in a composite cell construct for bone and cartilage  
CC regeneration. The polynucleotides can be employed to produce the  
CC polypeptides by recombinant techniques  
XX  
SQ Sequence 476 BP; 155 A; 106 C; 98 G; 117 T; 0 U; 0 Other;  
Query Match 52.9%; Score 40.2; DB 3; Length 476;  
Best Local Similarity 93.3%; Pred. No. 3.7e-05;  
Matches 42; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GATTTCCTTGGCCCTCTGGAGAGACCTTGTATGCCCT 45  
DB 59 GATCTGTTCCTGCTCTGGAGAGACCTTGTATGCCCT 103

RESULT 7  
ADB59026  
ID ADB59026 standard; DNA, 556 BP.  
XX  
AC ADB59026;  
XX  
DT 04-DEC-2003 (first entry)  
XX  
DE Toxicity-related gene, SEQ ID 4052.  
XX  
KW Toxic; toxin; gene expression profile; hepatotoxicity; liver;  
KW drug screening; toxicity assay; da.  
XX  
OS Unidentified.  
XX  
PN WO2003064624-A2.  
XX  
PD 07-AUG-2003.  
XX  
PF 31-JAN-2003; 2003WO-US003194.  
XX  
PR 31-JAN-2002; 2002US-00060087.  
PR 15-MAR-2002; 2002US-0364045P.  
PR 15-MAR-2002; 2002US-0364055P.  
PR 30-DEC-2002; 2002US-0436643P.  
XX  
PA (GENE-) GENE LOGIC INC.  
XX  
PI Mendrick D, Porter M, Johnson K, Higgs B, Castle A, Elashoff M;  
XX  
DR WPI; 2003-689530/65.  
XX  
PT Predicting a toxic effect of a compound, useful in identifying toxicity  
PT markers in liver tissues or cells for drug screening and toxicity assays,  
PT comprises preparing gene expression profile of tissue or cells exposed to

Pt			the compound.
Xx			
p8	Claim 1; SEQ ID NO 4052; 1156bp; English.		
Cc	The present invention relates to a method for predicting a toxic effect		
Cc	of a compound. The method comprises preparing a gene expression profile		
Cc	of a tissue or cell sample exposed to the compound, and comparing the		
Cc	gene expression profile to a database comprising SEQ ID 1-4925, where		
Cc	differential expression of the gene indicates at least one toxic effect.		
Cc	The method is useful for predicting at least one toxic effect of a		
Cc	compound, predicting hepatotoxicity or the progression of a toxic effect		
Cc	of a compound, identifying an agent that modulates the onset or		
Cc	progression of a toxic response, predicting the cellular pathways that a		
Cc	compound modulates in a cell, and identifying an agent that modulates at		
Cc	least one activity of a protein. The method and compositions of the		
Cc	present invention using a database of genes having liver toxin-induced		
Cc	differential expression, are useful in identifying toxicity markers in		
Cc	liver tissues or cells for drug screening and toxicity assays. Note: The		
Cc	sequence data for this patent did not form part of the printed		
Cc	specification, but was obtained in electronic format directly from WIPO		
Cc	at ftp.wipo.int/pub/published_pct_sequences.		
SQ			
	Sequence 556 BP; 121 A; 183 C; 114 G; 119 T; 0 U; 19 Other;		
	Query Match	51.6%; Score 39.2; DB 10; Length 556;	
	Best Local Similarity	91.1%; Pred. No. 9.5e-05;	
	Matches	41; Conservative 0; Mismatches 4; Indels 0; Gaps 0;	
Oy			
	1 GATTTTGTTTGCCTGCGCCTCTGGAGACGCCCTTGTCATAGCCCGT 45		
Dd			
	33 GATCTGTGTTGCTGCTCCCTCTGGAGACGCTTTGNTATNGCCCT 77		
RESULT 8			
AADBS3782			
ID	AADB53782 standard; DNA; 556 BP.		
XX			
AA	AADB53782;		
XC			
DD			
DT	04-DEC-2003 (first entry)		
DE	Primary rat hepatocyte toxicity modelling related gene SEQ ID NO:4324.		
KW	toxic effect; gene expression profile; hepatotoxicity; diagnostic marker;		
RW	toxicity marker; toxicity progression; drug screening;		
KV	primary rat hepatocyte toxicity modelling; gene; ds.		
XS			
OS	Rattus norvegicus.		
XX			
PN	WO2003065993-A2.		
XX			
PD	14-AUG-2003.		
PF			
XX	04-FEB-2003; 2003WO-US003462.		
XX			
PR	04-FEB-2002; 2002US-0353171P.		
PR	13-MAR-2002; 2002US-036534P.		
PR	08-APR-2002; 2002US-037024BP.		
PR	10-APR-2002; 2002US-0371134P.		
PR	10-APR-2002; 2002US-0371135P.		
PR	10-APR-2002; 2002US-0371150P.		
PR	11-APR-2002; 2002US-0371411P.		
PR	19-APR-2002; 2002US-0373601P.		
PR	19-APR-2002; 2002US-0373603P.		
PR	22-APR-2002; 2002US-0374139P.		
PR	08-MAY-2002; 2002US-0378370P.		
PR	09-MAY-2002; 2002US-0378652P.		
PR	09-MAY-2002; 2002US-0378653P.		
PR	09-JUL-2002; 2002US-0379665P.		
PR	09-JUL-2002; 2002US-0394230P.		
PR	04-SEP-2002; 2002US-0394253P.		
PR	04-SEP-2002; 2002US-0407688P.		
PR	28-JAN-2003; 2003US-0442900P.		

XX	(GENE-) GENE LOGIC INC.
PA	Mendrick D,
P1	Porter M, Johnson K, Higgs B, Caetle A, Orr M,
P1	Elaaboff M,
XX	WPI, 2003-731472/69.
DR	
XX	Determining if a compound induces a toxic effect on a tissue or cell, for
PT	identifying hepatotoxic compounds, comprises comparing a gene expression
PT	profile of a tissue or cell sample to a database of Tox mean and non-Tox
PT	mean values.
XX	
PS	Claim 44; SEQ ID NO 4324; 874bp; English.
XX	
CC	The present invention describes a method for determining whether a
CC	compound induces a toxic effect on a tissue or cell. The method comprises
CC	preparing a gene expression profile of a tissue or cell sample exposed to
CC	the compound, and comparing the gene expression profile to a database
CC	comprising data or information on the Tox mean and non-Tox mean value.
CC	The method is useful for predicting or identifying at least one toxic
CC	effect, particularly hepatotoxicity, of a test or unknown compound. The
CC	genes listed in the specification are useful as diagnostic or toxicity
CC	markers for the prediction or identification of the physiological state
CC	of tissue or cell sample that has been exposed to a compound, or to
CC	identify or predict the toxic effects of a compound or an agent. These
CC	may also be used as markers for monitoring toxicity progression or for
CC	drug screening. The present sequence represents a primary rat hepatocyte
CC	toxicity modelling related gene sequence from the present invention.
XX	
SQ	Sequence 556 BP; 121 A; 183 C; 114 G; 119 T; 0 U; 19 Other;
Query Match	51.6%; Score 39.2; DB 10; Length 556;
Best Local Similarity	91.1%; Pred. No. 9.5e-05;
Matches	41; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
OY	1 GAATTTGTTGGCTGCCTCCTCGAGACAGCCTTGTACTGCCGT 45
Db	33 GATCTTGTTGGCTGCCTCCTCGAGACAGCCTTGTATGCCCT 77
RESULT 9	
AAL41111	
ID	AAL41111 standard; DNA; 722 BP.
XX	
AC	AAL41111:
XX	
DT	16-OCT-2002 (first entry)
XX	
DE	gAML related Y-chromosome DNA sequence.
XX	
KM	Goat embryo sexual identification technique; goat amelogenin gene; gAML;
KW	sex-specific; gene; de; Y-chromosome.
XX	
OS	Capra hircus.
XX	
FT	Key Location/Qualifiers
FT	CDS 35..658
FT	/tag= a
FT	/product= "Y-chromosome protein"
XX	
PN	TW454013-A.
PD	11-SEP-2001.
XX	
PF	10-NOV-1999; 99TW-00119616.
XX	
PR	10-NOV-1999; 99TW-00119616.
XX	
PA	(CHEN/) CHEN C.
PA	(JANG/) JANG J.
PA	(WENG/) WENG T.
PA	(JENG/) JENG D.

XX	Chen C,	Jang J,	Weng T,	Jeng D;
PI				
XX	WPI,	2002-442016/47.		
DR	P-PsDB;	AAO22534.		
XX				
PT	Sex-specific sequence of goat amelogenin gene, useful for embryo sexual			
PT	identification, comprises high sensitivity even using single white blood			
PT	cell or cleavage c.			
XX				
PS	Disclosure; Page 28, 35pp; Chinese.			
XX				
CC	The invention relates to a goat embryo sexual identification technique			
CC	with high efficiency, sensitivity and repeatability. This technique			
CC	involves separately cloning and sequencing the coding regions and the			
CC	introns of the goat amelogenin gene (GAML) on the goat chromosomes. The			
CC	results indicate that there are sex-specific sequences in the fifth			
CC	intron of the gene. The major characteristics according to the present			
CC	invention include high sensitivity, applicable in sex identification even			
CC	without using a single white blood cell or a single cleavage cell of			
CC	blastula; high diagnostic efficiency, capable of identifying hundreds of			
CC	goat embryos in 3 hours; simple operation procedures without complicated			
CC	steps of DNA extraction and need no additional control group intron; and			
CC	can be applied on different species of goats. This polynucleotide			
CC	sequence represents a GAML related Y-chromosome DNA sequence of the			
CC	invention			
SQ				
	Sequence 722 BP; 173 A; 264 C; 140 G; 145 T; 0 U; 0 Other;			
	Query Match	51.1%;	Score 38.8;	DB 6;
	Best Local Similarity	95.2%;	Pred. No. 0.00015;	Length 722;
	Matches 40;	Conservative 0;	Mismatches 2;	Indels 0;
				Gaps 0;
QY	1	GATTTTGTTCCTCCTCGTGAGACGCTTGTCTAAGCC	42	
Db	46	GATTITGTTGCCTCCTCGTGAGACGCTTCTCTATGCC	87	
	RESULT 10			
	AAL41110			
ID	AAL41110 standard; DNA; 752 BP.			
XX				
AC	AAL41110;			
XX				
DT	16-OCT-2002 (first entry)			
XX				
DE	GAML related X-chromosome DNA sequence.			
XX				
KM	Goat embryo sexual identification technique; goat amelogenin gene; GAML;			
KV	sex-specific; gene; ds; X-chromosome.			
OS				
XX	Capra hircus.			
XX				
FH	Key	Location/Qualifiers		
FT	CDS	35..658		
FT		/tag= a		
FT		/product= "X-chromosome protein"		
PN	TW454013-A.			
PD	11-SEP-2001.			
PF	10-NOV-1999;	99TW-00119616.		
PR	10-NOV-1999;	99TW-00119616.		
PA	(CHEN//) CHEN C.			
PA	(JANG//) JANG J.			
PA	(WENG//) WENG T.			
PA	(JENG//) JENG D.			
XI	Chen C,	Jang J,	Weng T,	Jeng D;
XX				

DR	MP1: 2002-442016/47.
DR	P-PsDB; AAO22534.
XX	
PT	Sex-specific sequence of goat amelogenin gene, useful for embryo sexual
PT	identification, comprises high sensitivity even using single white blood
PT	cell or cleavage c.
XX	
PS	Disclosure; Page 28; 35pp; Chinese.
XX	
CC	The invention relates to a goat embryo sexual identification technique
CC	with high efficiency, sensitivity and repeatability. This technique
CC	involves separately cloning and sequencing the coding regions and the
CC	introns of the goat amelogenin gene (GAML) on the goat chromosomes. The
CC	results indicate that there are sex-specific sequences in the fifth
CC	intron of the gene. The major characteristics according to the present
CC	invention include high sensitivity, applicable in sex identification even
CC	only using a single white blood cell or a single cleavage cell of
CC	blactaria; high diagnostic efficiency, capable of identifying hundreds of
CC	goat embryo in 3 hours; simple operation procedures without complicated
CC	steps of DNA extraction and need no additional control group intron; and
CC	can be applied on different species of goats. This polynucleotide
CC	sequence represents a GAML related X-chromosome DNA sequence of the
CC	invention
XX	
SQ	Sequence 752 BP, 184 A, 271 C, 152 G, 145 T, 0 U; 0 Other;
	Query Match 51.1%; Score 38.8; DB 6; Length 752;
	Best Local Similarity 95.2%; Pred. NO. 0.00015;
	Matches 40; Conservative 0; Mismatches 2; Indels 0; Gaps 0
Qy	1 GATTTTGTTCCTGCTCCCTCGGAGACGCTTCTCTATGCC 42
Db	46 GATTTTGTTCCTGCTCCCTCGGAGAGCCTTCTCTATGCC 87
RESULT 11	
ADM80836	
ID	ADM80836 standard; cDNA, 318 BP.
XX	
AC	ADM80836;
XX	
DT	03-JUN-2004 (first entry)
XX	
DE	Human CADEC-M-23 encoding cDNA SEQ ID NO:65.
XX	
KW	human; cell adhesion and extracellular matrix protein; CADEC-M;
KW	neuroprotective; cytoskeletal; anorectic; immune disorder;
KW	neurological disorder; developmental disorder;
KW	connective tissue disorder; cell proliferative disorder; cancer; obesity;
KW	Tangier disease; gene; ss.
XX	
XX	Homo sapiens.
OS	
XX	
XX	Key Location/Qualifiers
FT	CDS 36..260
FT	/*tag= a
FT	/product= "CADEC-M-23"
XX	
XX	WO2004015396-A2.
XX	
XX	19-FEB-2004.
PD	
PF	12-AUG-2003; 2003MO-US025418.
XX	
PR	13-AUG-2002; 2002US-0403781P.
PR	30-AUG-2002; 2002US-0407034P.
PR	13-SEP-2002; 2002US-0410566P.
PR	24-SEP-2002; 2002US-0413482P.
PR	25-SEP-2002; 2002US-0413890P.
PR	08-NOV-2002; 2002US-0424904P.
PR	13-NOV-2002; 2002US-0426222P.
XX	
XX	(INCY- ) INCYTE CORP.

XX	Elliott VS., Khare R., Emerling BM., Kable AE., Tran UK., Jin P.
P1	Becha SD., Margulis JP., Swarnaker A., Chavla NK., Ramkumar J.
P1	Hatalia AJL., Lee SY., Jiang X., Jackson AA., Richardson TW., Blake JJ.
P1	Wang JT., Chien D., Yang YG.
XX	
XX	WPI; 2004-191795/18.
DR	P-PsDB; ADM80794.
PT	New cell adhesion and extracellular matrix proteins, useful in
PT	diagnosing, treating and preventing immune, neurological, developmental,
CC	connective tissue and cell proliferative disorders including cancer.
PS	Claim 5; SEQ ID NO 65; 272bp; English.
XX	
CC	The present sequence encodes a human cell adhesion and extracellular
CC	matrix protein designated CADSCM. CADSCM sequences has neuroprotective,
CC	cystostatic and antiretic activities. The CADSCM polypeptides and
CC	polynucleotides are useful in diagnosing, treating and preventing immune,
CC	neurological, developmental, connective tissue and cell proliferative
CC	disorders including cancer, e.g. breast, prostate, ovarian, lung or colon
CC	cancer, obesity and Tangier disease.
SQ	Sequence 318 BP; 87 A; 81 C; 75 G; 75 T; 0 U; 0 Other;
	Query Match            48.9%; Score 37.2; DB 12; Length 318;
	Best Local Similarity   92.9%; Pred. No. 0.00047;
	Matches     39; Conservative     0; Mismatches     3; Indels     0; Gaps     0
Oy	1 GATTTCGTTGGCTGCCTCCTCGAGACAGCCTTGTATGCC 42
Dd	47 GATTTCGTTGGCTGCCTCCTCGAGACAGCCTTGTATGCC 88
RESULT 12	
AAZ07020	
ID	AAZ07020 standard; DNA; 750 BP.
XX	
AC	AAZ07020;
XX	
DT	15-NOV-1999 (first entry)
DE	Amelogenin X nucleotide sequence.
XX	
KW	Amelogenin X; Amelogenin Y; genotype; 5' nuclease amplification;
KW	multiple allelic site; apolipoprotein E; apoB; coronary artery disease;
KW	Alzheimer's disease; ds.
OS	Unidentified.
XX	
PN	WO9940226-A2.
PD	12-AUG-1999.
XX	
PF	08-JAN-1999; 99WO-US000499.
XX	
PR	04-FEB-1998; 98US-00018595.
XX	
PA	(PEKE ) PERKIN-ELMER CORP.
P1	Livak KJ, Goodsaal F;
XX	
XX	WPI; 1999-539985/45.
DR	
PT	5' nuclease amplification assay using fluorescence-quencher probes for
PT	determination of a genotype at multiple allelic sites.
XX	
XX	Disclosure; Fig 10; 95pp; English.
CC	
CC	The present invention describes first and second sets of fluorescer-
CC	quencher probes used simultaneously in a 5' nuclease assay to identify
CC	which members of a first or second set of substantially homologous
CC	sequences are present in a DNA sample. The method can be used to genotype

CC	a sample of genomic DNA at two or more different allelic sites.
CC	Generating a fluorescence spectrum and signature for each genotype, which
CC	uniquely reflects the assay's inherent inefficiency for that genotype
CC	given the particular conditions, probes and primers used, the genotype of
CC	unknown sequences can be determined. The assay was shown to be useful for
CC	determining apoe genotypes. The assay can be used as a diagnostic tool
CC	for assessing the risk for coronary artery disease and/or late-onset
CC	Alzheimer's disease. Using the 5' nuclease assay of the invention it is
CC	possible to determine a genotype at two or more allelic sites in a single
CC	reaction. This approach is much faster than previous approaches to
CC	genotyping genes having two or more allelic sites, such as the
CC	apolipoprotein E gene. A key advantage of the method for determining the
CC	genotype of a sample of DNA at multiple allelic sites is that it does not
CC	rely on 5' nuclease assay working with 100% efficiency to distinguish
CC	between substantially homologous sequences such as alleles. The present
CC	sequence represent the nucleotide sequence for amelogenin X, which is
CC	used in the exemplification of the present invention
XX	
SQ	Sequence 750 BP; 191 A; 260 C; 146 G; 153 T; 0 U; 0 Other;
Query Match	48.9%; Score 37.2; DB 2; Length 750;
Best Local Similarity	92.9%; Pred. No. 0.0062;
Matches 39; Conservative	0; Mismatches 3; Indels 0; Gaps 0;
Oy	1 GATTTTGTTGGCTTCGCTCCTCGAGACAGCCTTGATGCC 42 
Db	80 GATTTTATTGGCCCTGCCTCCTCGAGACAGCCTTTGGCATGCC 121 
RESULT 13	
AAZ07018	
ID	AAZ07018 standard; DNA; 793 BP.
XX	
AC	AAZ07018;
DT	15-NOV-1999 (first entry)
XX	
DE	Amelogenin X nucleotide sequence.
XX	
KW	Amelogenin X; Amelogenin Y; genotype; 5' nuclease amplification; multiple allelic site; apolipoprotein E; apoe; coronary artery disease; Alzheimer's disease; se.
XX	
OS	Unidentified.
XX	
PN	MO9940226-A2.
XX	
PD	12-AUG-1999.
XX	
PF	08-JAN-1999; 99WO-US000499.
XX	
PR	04-FEB-1998; 98US-00018595.
XX	
PA	(PEKE ) PERKIN-ELMER CORP.
XX	
P1	Livak KJ, Goodaaid F;
XX	
DR	WPI; 1999-539985/45.
XX	
PT	5' nuclease amplification assay using fluorescence-quencher probes for
XX	determination of a genotype at multiple allelic sites.
XX	
XX	Disclosure; Fig 8A; 95pp; English.
CC	The present invention describes first and second sets of fluorescer-
CC	-quencher probes used simultaneously in a 5' nuclease assay to identify
CC	certain members of a first or second set of substantially homologous
CC	sequences are present in a DNA sample. The method can be used to genotype
CC	a sample of genomic DNA at two or more different allelic sites.
CC	Generating a fluorescence spectrum and signature for each genotype, which
CC	uniquely reflects the assay's inherent inefficiency for that genotype
CC	given the particular conditions, probes and primers used, the genotype of
CC	unknown sequences can be determined. The assay was shown to be useful for

CC determining apoe genotypes. The assay can be used as a diagnostic tool  
 CC for assessing the risk for coronary artery disease and/or late-onset  
 CC Alzheimer's disease. Using the 5' nuclease assay of the invention it is  
 CC possible to determine a genotype at two or more allelic sites in a single  
 CC reaction. This approach is much faster than previous approaches to  
 CC genotyping genes having two or more allelic sites, such as the  
 CC apolipoprotein E gene. A key advantage of the method for determining the  
 CC genotype of a sample of DNA at multiple allelic sites is that it does not  
 CC rely on 5' nuclease assay working with 100% efficiency to distinguish  
 CC between substantially homologous sequences such as alleles. The present  
 CC sequence represent the nucleotide sequence for ameloigenin X, which is  
 CC used in the exemplification of the present invention

XX Sequence 793 BP; 215 A; 266 C; 150 G; 162 T; 0 U; 0 Other;

Query Match 48.9%; Score 37.2; DB 2; Length 793;  
 Best Local Similarity 92.9%; Pred. No. 0.0063;  
 Matches 39; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GATTTGTTGCTGCTGCTGCGAGCAGCTTTGCTATGCC 42  
 DB 80 GATTTATTTGCTGCTGCTGCGAGCAGCTTTGCTATGCC 121

RESULT 14  
 AA207019  
 ID AA207019 standard; DNA; 802 BP.

XX AA207019;

DT 15-NOV-1999 (first entry)

XX Ameloigenin Y nucleotide sequence.

XX Ameloigenin X; Ameloigenin Y; genotype; 5' nuclease amplification;  
 KW multiple allelic site; apolipoprotein E; apoe; coronary artery disease;  
 KM Alzheimer's disease; 89.

XX Unidentified.

OS WO9940226-A2.

PN 12-AUG-1999.

PD 08-JAN-1999; 99WO-US000499.

PF 04-FEB-1998; 98US-00018595.

PR (PERKE) PERKIN-ELMER CORP.

PA Llvak KJ, Goodaaid F;

PI WPI; 1999-539985/45.

DR 5' nuclease amplification assay using fluorescence-quencher probes for  
 XX determination of a genotype at multiple allelic sites.

PT Disclosure; Fig 8B; 95pp; English.

XX The present invention describes first and second sets of fluorescer-  
 CC quencher probes used simultaneously in a 5' nuclease assay to identify  
 CC which members of a first or second set of substantially homologous  
 CC sequences are present in a DNA sample. The method can be used to genotype  
 CC a sample of genomic DNA at two or more different allelic sites.  
 CC Generating a fluorescence spectrum and signature for each genotype, which  
 CC uniquely reflects the assay's inherent inefficiency for that genotype  
 CC given the particular conditions, probes and primers used, the genotype of  
 CC unknown sequences can be determined. The assay was shown to be useful for  
 CC determining apoe genotypes. The assay can be used as a diagnostic tool  
 CC for assessing the risk for coronary artery disease and/or late-onset  
 CC Alzheimer's disease. Using the 5' nuclease assay of the invention it is  
 CC possible to determine a genotype at two or more allelic sites in a single  
 CC reaction. This approach is much faster than previous approaches to

CC genotyping genes having two or more allelic sites, such as the  
 CC apolipoprotein E gene. A key advantage of the method for determining the  
 CC genotype of a sample of DNA at multiple allelic sites is that it does not  
 CC rely on 5' nuclease assay working with 100% efficiency to distinguish  
 CC between substantially homologous sequences such as alleles. The present  
 CC sequence represent the nucleotide sequence for ameloigenin Y, which is  
 CC used in the exemplification of the present invention

XX Sequence 802 BP; 216 A; 258 C; 155 G; 173 T; 0 U; 0 Other;

Query Match 46.8%; Score 35.6; DB 2; Length 802;  
 Best Local Similarity 90.5%; Pred. No. 0.0026;  
 Matches 38; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GATTTGTTGCTGCTGCTGCGAGCAGCTTTGCTATGCC 42  
 DB 80 GATTTGTTGCTGCTGCTGCGAGCAGCTTTGCTATGCC 121

RESULT 15  
 ADQ22977  
 ID ADQ22977 standard; DNA; 852 BP.

XX ADQ22977;

DT 26-AUG-2004 (first entry)

XX Human soft tissue sarcoma-upregulated DNA - SEQ ID 5797.

XX soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;

XX Homo sapiens.

OS WO2004048938-A2.

PN 10-JUN-2004.

PD 26-NOV-2003; 2003WO-US038193.

PF 26-NOV-2002; 2002US-0429739P.

PR (PROT-) PROTEIN DESIGN LABS INC.

PA Aziz N, Gainsburg WM, Zlotnik A;

PI WPI; 2004-441208/41.

DR Early detection of soft tissue sarcoma comprises determining expression

XX of a gene in a first soft tissue sample and a normal soft tissue sample

PT and comparing the gene expression, also useful in treating soft tissue

PT sarcoma.  
 XX Example 2; SEQ ID NO 5797; 210pp; English.

XX The invention relates to a novel method for detecting soft tissue sarcoma  
 CC which comprises obtaining a first soft tissue sample from an individual,  
 CC and a normal soft tissue sample from the same or different individual,  
 CC determining the expression of a gene in both samples and comparing the  
 CC expression of the gene in both soft tissue samples, where a higher level  
 CC of protein expression in the first soft tissue sample indicates the  
 CC presence of soft tissue sarcoma. The method of the invention has  
 CC cytostatic applications and may be useful for detecting soft tissue  
 CC sarcoma, possibly via gene therapy or vaccine production. The nucleic  
 CC acid sequences may be useful in diagnostic and screening applications.  
 CC The current sequence is that of a human soft tissue sarcoma-upregulated  
 CC DNA of the invention. The current sequence is not shown within the  
 CC specification per se but was submitted in CD format by the inventor.

XX Sequence 852 BP; 226 A; 267 C; 169 G; 190 T; 0 U; 0 Other;

Query Match 46.8%; Score 35.6; DB 12; Length 852;  
 Best Local Similarity 90.5%; Pred. No. 0.0027;

Wed Feb 9 09:28:18 2005

us-10-754-437-6.rng

**Page 9**

<b>Matches</b>	<b>38;</b>	<b>Conservative</b>	<b>0;</b>	<b>Mismatches</b>	<b>4;</b>	<b>Indels</b>	<b>0;</b>	<b>Gaps</b>	<b>0;</b>
----------------	------------	---------------------	-----------	-------------------	-----------	---------------	-----------	-------------	-----------

  

<b>Oy</b>	<b>1</b>	<b>GATTTGTTTGCTGCCTCCCGAGAGCAGCCCTTGGCTATGCC</b>	<b>42</b>
<b>Db</b>	<b>80</b>	<b>GATTTGTGTTGCTGCCTTGGGAGAGCACTTTGGCCATGCC</b>	<b>121</b>

```
Search completed: February  8, 2005, 13:57:00
Job time : 104.817 secs
```

**This Page Blank (uspto)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 8, 2005, 13:03:55 ; Search time 20.5262 Seconds  
(without alignments)  
2631.757 Million cell updates/sec

Title: US-10-754-437-6

Perfect score: 76  
Sequence: 1 gattctgttcgtcgtcctcc.....ccctgcataataatcaattt 76

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 824507 seqs, 355394441 residues

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents NA:  
1: /cgn2\_6/prodata/1/ina/5A COMB.seq.\*  
2: /cgn2\_6/prodata/1/ina/5B COMB.seq.\*  
3: /cgn2\_6/prodata/1/ina/6A COMB.seq.\*  
4: /cgn2\_6/prodata/1/ina/6B COMB.seq.\*  
5: /cgn2\_6/prodata/1/ina/PCUS COMB.seq.\*  
6: /cgn2\_6/prodata/1/ina/backfile1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	40.2	52.9	476	4 US-09-744-128-3	Sequence 3, Appl 1
2	37.2	48.9	500	2 US-09-018-595B-3	Sequence 3, Appl 1
3	37.2	48.9	500	3 US-09-324-709A-3	Sequence 3, Appl 1
4	37.2	48.9	793	2 US-09-018-595B-1	Sequence 1, Appl 1
5	37.2	48.9	793	3 US-09-324-709A-1	Sequence 1, Appl 1
6	35.6	46.8	802	2 US-09-018-595B-2	Sequence 2, Appl 1
7	35.6	46.8	802	3 US-09-324-709A-2	Sequence 2, Appl 1
8	25.6	33.7	936	4 US-09-489-039A-6691	Sequence 6691, Ap
9	25.2	33.2	274	4 US-09-313-294A-3198	Sequence 3198, Ap
10	25.2	33.2	277	4 US-09-313-294A-5716	Sequence 5716, Ap
11	24.2	31.8	930	4 US-09-710-279-1607	Sequence 1607, Ap
12	24.2	31.8	1311	3 US-09-134-001C-386	Sequence 386, App
13	24.2	31.8	1315	4 US-09-710-279-1381	Sequence 3881, App
14	24	31.6	1107	4 US-09-252-991A-4254	Sequence 4254, Ap
15	23.6	31.1	576	4 US-09-489-039A-4185	Sequence 4185, Ap
16	23.4	30.8	684	4 US-09-270-767-7458	Sequence 7458, Ap
17	23.4	30.8	684	4 US-09-270-767-22740	Sequence 22740, A
18	23.4	30.8	1664576	4 US-08-916-421B-1	Sequence 1, Appl 1
19	23.4	30.8	1664976	4 US-09-692-570-1	Sequence 1, Appl 1
20	23	30.3	1386	4 US-09-543-681A-2276	Sequence 2276, Ap
21	22.8	30.0	826	6 US-09-189-760-5	Patent No. 5252477
22	22.8	30.0	1529	3 US-09-188-811-5	Sequence 5, Appl 1
23	22.8	30.0	1529	3 US-09-188-811-5	Sequence 5, Appl 1
24	22.8	30.0	1529	3 US-09-514-422-5	Sequence 5, Appl 1
25	22.8	30.0	2791	4 US-09-799-451-500	Sequence 500, App
26	22.6	29.7	461	4 US-09-513-999C-1419	Sequence 1419, Ap
27	22.6	29.7	1080	6 US-09-223-3	Patent No. 5260223

## ALIGNMENTS

C 28	22.6	29.7	2039	4	US-09-799-451-583	Sequence 583, App
C 29	22.6	29.7	2087	4	US-09-799-451-582	Sequence 582, App
C 30	22.6	29.7	1830121	4	US-09-557-884-1	Sequence 1, Appl 1
C 31	22.6	29.7	1830121	4	US-09-643-990A-1	Sequence 1, Appl 1
C 32	22.6	29.7	1830121	4	US-10-329-960-1	Sequence 1, Appl 1
C 33	22.4	29.5	561	4	US-09-248-796A-14014	Sequence 14014, A
C 34	22.4	29.5	681	4	US-09-134-000C-2467	Sequence 2467, Ap
C 35	22.4	29.5	3860	4	US-09-620-312D-873	Sequence 873, App
C 36	22.4	29.5	7195	3	US-08-478-507-6	Sequence 6, Appl 1
C 37	22.4	29.5	7195	3	US-09-128-275A-6	Sequence 6, Appl 1
C 38	22.4	29.5	7195	3	US-09-553-427-6	Sequence 6, Appl 1
C 39	22.4	29.5	14636	3	US-09-173-914-6	Sequence 6, Appl 1
C 40	22.4	29.5	786431	4	US-09-751-389-3	Sequence 3, Appl 1
C 41	22.2	29.2	11933	3	US-09-470-618-13	Sequence 13, Appl 1
C 42	22.2	29.2	11933	3	US-09-364-862-13	Sequence 13, Appl 1
C 43	22.2	29.2	161652	4	US-09-497-855A-40	Sequence 40, Appl 1
C 44	22	28.9	592	3	US-08-896-164-80	Sequence 80, Appl 1
C 45	22	28.9	667	4	US-09-221-017B-597	Sequence 597, App

RESULT 1  
US-09-744-128-3  
Sequence 3, Application US/09744128  
Patent No. 6677306  
GENERAL INFORMATION:  
APPLICANT: Vels et al.  
TITLE OF INVENTION: Chondrogenic and Osteogenic Inducing Molecule  
FILE REFERENCE: 27636/36983  
CURRENT APPLICATION NUMBER: US/09/744,128  
PRIOR FILING DATE: 2001-05-16  
PRIOR APPLICATION NUMBER: PCT/US99/17342  
PRIOR FILING DATE: 1999-07-29  
PRIOR APPLICATION NUMBER: 60/094,489  
NUMBER OF SEQ ID NOS: 17  
SOFTWARE: PatentIn 3.1  
SEQ ID NO 3  
LENGTH: 476  
TYPE: DNA  
ORGANISM: Artificial sequence  
FEATURES:  
OTHER INFORMATION: Description of Artificial Sequence: PCR Product  
US-09-744-128-3  
Query Match 52.9%; Score 40.2; DB 4; Length 476;  
Best Local Similarity 93.3%; Pred. No. 2.6e-06;  
Matches 42; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 GATTTGTTGCTGCTGCTGCGAGCAGCTTGTATGCCCGT 45  
59 GATCTGTGCTGCTGCTGCTGCGAGCAGCTTGTATGCCCGT 103  
RESULT 2  
US-09-018-595B-3  
Sequence 3, Application US/09018595B  
Patent No. 596223  
GENERAL INFORMATION:  
APPLICANT: Perkin-Elmer Corporation,  
APPLICANT: Applied Biosystems Division  
TITLE OF INVENTION: DETERMINATION OF GENOTYPE OF  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSER: David J. Weitz,  
ADDRESSER: Wilson Somsini Goodrich & Rosati  
STREET: 650 Page Mill Road  
CITY: Palo Alto  
STATE: California  
COUNTRY: USA

ZIP: 94304-1050  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch diskette  
COMPUTER: IBM compatible  
OPERATING SYSTEM: Microsoft Windows 95  
SOFTWARE: Wordperfect for windows 6.0,  
SOFTWARE: ASCII (DOS) TEXT format  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/018,595B  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
ATTORNEY/AGENT INFORMATION:  
NAME: David J. Weitz  
REGISTRATION NUMBER: 38,362  
REFERENCE/DOCKET NUMBER: PELM-744  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 493-9300  
TELEFAX: (650) 493-6811  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 500 nucleotides  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-09-018-595B-3

Query Match 48.9%; Score 37.2; DB 2; Length 500;  
Best Local Similarity 92.9%; Pred. No. 4,1e-05;  
Matches 39; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GATTTGTTGCTGCTCTCTGGAGACGCTTTGCTATGCC 42  
DB 80 GATTTATTTGCTGCTCTCTGGAGACGCTTTGCTATGCC 121

RESULT 3  
US-09-324-709A-3  
Sequence 3, Application US/09324709A  
Patent No. 6134707  
GENERAL INFORMATION:  
APPLICANT: Perkin-Elmer Corporation,  
APPLICANT: Applied Biosystems Division  
TITLE OF INVENTION: FLUORESCENCE GENOTYPING AT MULTIPLE ALLELIC  
TITLE OF INVENTION: SITES  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David J. Weitz,  
ADDRESSEE: Wilson Sonsini Goodrich & Rosati  
STREET: 650 Page Mill Road  
CITY: Palo Alto  
STATE: California  
COUNTRY: USA  
ZIP: 94304-1050  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch diskette  
COMPUTER: IBM compatible  
OPERATING SYSTEM: Microsoft Windows 95  
SOFTWARE: Wordperfect for windows 6.0,  
SOFTWARE: ASCII (DOS) TEXT format  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/324,709A  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
ATTORNEY/AGENT INFORMATION:  
NAME: David J. Weitz  
REGISTRATION NUMBER: 38,362  
REFERENCE/DOCKET NUMBER: 16842-758

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 493-9300  
TELEFAX: (650) 493-6811  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 500 nucleotides  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-09-324-709A-3

Query Match 48.9%; Score 37.2; DB 3; Length 500;  
Best Local Similarity 92.9%; Pred. No. 4,1e-05;  
Matches 39; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GATTTGTTGCTGCTCTCTGGAGACGCTTTGCTATGCC 42  
DB 80 GATTTATTTGCTGCTCTCTGGAGACGCTTTGCTATGCC 121

RESULT 4  
US-09-018-595B-1  
Sequence 1, Application US/09018595B  
Patent No. 5962233  
GENERAL INFORMATION:  
APPLICANT: Perkin-Elmer Corporation,  
APPLICANT: Applied Biosystems Division  
TITLE OF INVENTION: DETERMINATION OF GENOTYPE OF  
TITLE OF INVENTION: AMPLIFICATION PRODUCTS AT MULTIPLE ALLELIC SITES  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David J. Weitz,  
ADDRESSEE: Wilson Sonsini Goodrich & Rosati  
STREET: 650 Page Mill Road  
CITY: Palo Alto  
STATE: California  
COUNTRY: USA  
ZIP: 94304-1050  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch diskette  
COMPUTER: IBM compatible  
OPERATING SYSTEM: Microsoft Windows 95  
SOFTWARE: Wordperfect for windows 6.0,  
SOFTWARE: ASCII (DOS) TEXT format  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/018,595B  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
ATTORNEY/AGENT INFORMATION:  
NAME: David J. Weitz  
REGISTRATION NUMBER: 38,362  
REFERENCE/DOCKET NUMBER: PELM-744  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 493-9300  
TELEFAX: (650) 493-6811  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 793 nucleotides  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-018-595B-1

Query Match 48.9%; Score 37.2; DB 2; Length 793;  
Best Local Similarity 92.9%; Pred. No. 5e-05;  
Matches 39; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GATTTGTTGCTGCTCTCTGGAGACGCTTTGCTATGCC 42  
DB 80 GATTTATTTGCTGCTCTCTGGAGACGCTTTGCTATGCC 121

RESULT 5  
US-09-324-709A-1  
Sequence 1, Application US/09324709A  
Patent No. 6154707  
GENERAL INFORMATION:  
APPLICANT: Perkin-Elmer Corporation,  
APPLIED Biosystems Division  
TITLE OF INVENTION: FLUORESCENCE GENOTYPING AT MULTIPLE ALLELIC  
SITES  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David J. Weltz,  
STREET: 650 Page Mill Road  
CITY: Palo Alto  
STATE: California  
COUNTRY: USA  
ZIP: 94304-1050  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch diskette  
COMPUTER: IBM compatible  
OPERATING SYSTEM: Microsoft Windows 95  
SOFTWARE: Wordperfect for windows 6.0,  
SOFTWARE: ASCII (DOS) TEXT format  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/324,709A  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: David J. Weltz  
REGISTRATION NUMBER: 38,362  
REFERENCE/DOCKET NUMBER: 16842-758  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 493-9300  
TELEFAX: (650) 493-6811  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 793 nucleotides  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-324-709A-1

Query Match 48.9%; Score 37.2; DB 3; Length 793;  
Best Local Similarity 92.9%; Pred. No. 5e-05; 3; Indels 0; Gaps 0;  
Matches 39; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GATTTGTTGCTGCTGCTGCTGAGACAGCTTTGCTATGCC 42  
Db 80 GATTTATTTGCTGCTGCTGCTGAGACAGCTTTGCAATGCC 121

RESULT 6  
US-09-018-595B-2  
Sequence 2, Application US/09018595B  
Patent No. 5962233  
GENERAL INFORMATION:  
APPLICANT: Perkin-Elmer Corporation,  
APPLIED Biosystems Division  
TITLE OF INVENTION: DETERMINATION OF GENOTYPE OF  
TITLE OF INVENTION: AMPLIFICATION PRODUCTS AT MULTIPLE ALLELIC SITES  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David J. Weltz,  
STREET: 650 Page Mill Road  
CITY: Palo Alto  
STATE: California

COUNTRY: USA  
ZIP: 94304-1050  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch diskette  
COMPUTER: IBM compatible  
OPERATING SYSTEM: Microsoft Windows 95  
SOFTWARE: Wordperfect for windows 6.0,  
SOFTWARE: ASCII (DOS) TEXT format  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/018,595B  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: David J. Weltz  
REGISTRATION NUMBER: 38,362  
REFERENCE/DOCKET NUMBER: PELM-744  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 493-9300  
TELEFAX: (650) 493-6811  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 802 nucleotides  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-018-595B-2

Query Match 46.8%; Score 35.6; DB 2; Length 802;  
Best Local Similarity 90.5%; Pred. No. 0.00021;  
Matches 38; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 GATTTGTTGCTGCTGCTGCTGAGACAGCTTTGCTATGCC 42  
Db 80 GATTTGTTGCTGCTGCTGCTGAGACAGCTTTGCAATGCC 121

RESULT 7  
US-09-324-709A-2  
Sequence 2, Application US/09324709A  
Patent No. 6154707  
GENERAL INFORMATION:  
APPLICANT: Perkin-Elmer Corporation,  
APPLIED Biosystems Division  
TITLE OF INVENTION: FLUORESCENCE GENOTYPING AT MULTIPLE ALLELIC  
SITES  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David J. Weltz,  
STREET: 650 Page Mill Road  
CITY: Palo Alto  
STATE: California  
COUNTRY: USA  
ZIP: 94304-1050  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch diskette  
COMPUTER: IBM compatible  
OPERATING SYSTEM: Microsoft Windows 95  
SOFTWARE: Wordperfect for windows 6.0,  
SOFTWARE: ASCII (DOS) TEXT format  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/324,709A  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: David J. Weltz  
REGISTRATION NUMBER: 38,362

Query Match	31.8%;	Score 24.2;	DB 4;	Length 930;
Best Local Similarity	62.3%;	Pred. No. 7.1;		

Matches 38; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 10 TGGCTGCTCTCGGAGCAGCCTTTGCTATGCCCGTGAGTAAATACCTCGATATAT 69

DB 345 TGCATTTTTCATAGTCGACAGCCTTAATACGCCGGTAGTACATTACATTCATATATGT 404

QY 70 T 70

DB 405 T 405

RESULT 12  
US-09-134-001C-386

; Sequence 386, Application US/09134001C

; Patent No. 6380370

; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

; FILE REFERENCE: GTC-007

; CURRENT APPLICATION NUMBER: US/09/134,001C

; PRIOR FILING DATE: 1998-08-13

; PRIOR APPLICATION NUMBER: US 60/064,964

; PRIOR FILING DATE: 1997-11-08

; PRIOR APPLICATION NUMBER: US 60/055,779

; PRIOR FILING DATE: 1997-08-14

; NUMBER OF SEQ ID NOS: 5674

; SEQ ID NO 386

; LENGTH: 1311

; TYPE: DNA

; ORGANISM: Staphylococcus epidermidis

US-09-134-001C-386

Query Match

Best Local Similarity 31.8%; Score 24.2; DB 3; Length 1311;

Best Local Similarity 62.3%; Pred. No. 8.2;

Matches 38; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 10 TGGCTGCTCTCGGAGCAGCCTTTGCTATGCCCGTGAGTAAATACCTCGATATAT 69

DB 726 TGCATTTTTCATAGTCGACAGCCTTAATACGCCGGTAGTACATTACATTCATATATGT 785

QY 70 T 70

DB 786 T 786

RESULT 13

US-09-710-279-3881/c

; Sequence 3881, Application US/09710279

; Patent No. 6703492

; GENERAL INFORMATION:

; APPLICANT: KIMBERLY, WILLIAM JOHN

; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS

; FILE REFERENCE: PU3480US

; CURRENT APPLICATION NUMBER: US/09/710,279

; PRIOR FILING DATE: 2000-11-09

; PRIOR APPLICATION NUMBER: 60/164,258

; PRIOR FILING DATE: 1999-11-09

; NUMBER OF SEQ ID NOS: 4472

; SOFTWARE: Patentin Ver. 2.1

; SEQ ID NO 3881

; LENGTH: 3155

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE: OTHER INFORMATION: Description of Artificial Sequence: synthetic

; OTHER INFORMATION: nucleic acid sequence

US-09-710-279-3881

Query Match

Best Local Similarity 31.8%; Score 24.2; DB 4; Length 3155;

Best Local Similarity 62.3%; Pred. No. 12;

Matches 38; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 10 TGGCTGCTCTCGGAGCAGCCTTTGCTATGCCCGTGAGTAAATACCTCGATATAT 69

DB 2712 TGCATTTTTCATAGTCGACAGCCTTAATACGCCGGTAGTACATTACATTCATATATGT 2653

QY 70 T 70

DB 2652 T 2652

RESULT 14  
US-09-252-991A-4254/c

; Sequence 4254, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; PRIOR FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 4254

; LENGTH: 1107

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-4254

Query Match

Best Local Similarity 31.6%; Score 24; DB 4; Length 1107;

Best Local Similarity 75.0%; Pred. No. 9.2;

Matches 30; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 6 TGGTTCCTGCTCTCGGAGCAGCCTTTGCTATGCCCGT 45

DB 1104 TTTTTCCTTCCTCTCTGGAAGGCGCATTTGCTTGTCTGT 1065

RESULT 15

US-09-489-039A-4185/c

; Sequence 4185, Application US/09489039A

; Patent No. 6610836

; GENERAL INFORMATION:

; APPLICANT: Gary Breton et. al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA

; FILE REFERENCE: 2709.2004001

; CURRENT APPLICATION NUMBER: US/09/489,039A

; PRIOR FILING DATE: 2000-01-27

; PRIOR APPLICATION NUMBER: US 60/117,747

; PRIOR FILING DATE: 1999-01-29

; NUMBER OF SEQ ID NOS: 14342

; SEQ ID NO 4185

; LENGTH: 576

; TYPE: DNA

; ORGANISM: Klebsiella pneumoniae

US-09-489-039A-4185

Query Match

Best Local Similarity 31.1%; Score 23.6; DB 4; Length 576;

Best Local Similarity 61.3%; Pred. No. 10;

Matches 38; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 14 TGGCTGCTCTCGGAGCAGCCTTTGCTATGCCCGTGAGTAAATACCTCGATATATTTCA 73

DB 540 TGGTTCCTGCTCTCGGAGCAGCCTTTGCTATGCCCGTGAGTAAATACCTCGATATATTTCA 481

QY 74 TT 75

DB 480 TT 479

Wed Feb 9 09:28:19 2005

us-10-754-437-6.rni

Page 6

Search completed: February 8, 2005, 22:26:03  
Job time : 26.6373 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using SW model

Run on: February 8, 2005, 16:37:21 ; Search time 102.41 Seconds  
(without alignments)  
4264.119 Million cell updates/sec

Title: US-10-754-437-6

Perfect score: 76  
Sequence: 1 gattctgttcgtcctcctcc.....ccctgcataatcatcattt 76

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4300275 seqs, 2872944193 residues

Total number of hits satisfying chosen parameters: 8600550

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:\*

- 1: /cgn2\_6/prodata/2/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/prodata/2/pubpna/PC7\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/prodata/2/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/prodata/2/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/prodata/2/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/prodata/2/pubpna/PC7US\_PUBCOMB.seq:\*
- 7: /cgn2\_6/prodata/2/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/prodata/2/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/prodata/2/pubpna/US09A\_PUBCOMB.seq:\*
- 10: /cgn2\_6/prodata/2/pubpna/US09B\_PUBCOMB.seq:\*
- 11: /cgn2\_6/prodata/2/pubpna/US09C\_PUBCOMB.seq:\*
- 12: /cgn2\_6/prodata/2/pubpna/US09D\_PUBCOMB.seq:\*
- 13: /cgn2\_6/prodata/2/pubpna/US10A\_PUBCOMB.seq:\*
- 14: /cgn2\_6/prodata/2/pubpna/US10B\_PUBCOMB.seq:\*
- 15: /cgn2\_6/prodata/2/pubpna/US10C\_PUBCOMB.seq:\*
- 16: /cgn2\_6/prodata/2/pubpna/US10D\_PUBCOMB.seq:\*
- 17: /cgn2\_6/prodata/2/pubpna/US10E\_PUBCOMB.seq:\*
- 18: /cgn2\_6/prodata/2/pubpna/US10F\_PUBCOMB.seq:\*
- 19: /cgn2\_6/prodata/2/pubpna/US10G\_PUBCOMB.seq:\*
- 20: /cgn2\_6/prodata/2/pubpna/US10H\_PUBCOMB.seq:\*
- 21: /cgn2\_6/prodata/2/pubpna/US10I\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	52	68.4	600	US-10-027-632-287049	Sequence 287049,
C 2	52	68.4	600	US-10-027-632-287050	Sequence 287050,
C 3	52	68.4	600	US-10-027-632-287049	Sequence 287049,
C 4	52	68.4	600	US-10-027-632-287050	Sequence 287050,
C 5	52	68.4	259202	US-10-723-860-1311	Sequence 1311, Ap
C 6	40.2	52.9	65	US-09-908-975-1137	Sequence 1137, Ap
C 7	37.2	48.9	500	US-10-104-774-3	Sequence 3, Appli
C 8	37.2	48.9	500	US-10-455-150-3	Sequence 3, Appli
C 9	37.2	48.9	793	US-10-104-774-1	Sequence 1, Appli
C 10	37.2	48.9	793	US-10-455-150-1	Sequence 1, Appli
C 11	37.2	48.9	793	US-10-755-889-605	Sequence 605, App
C 12	35.6	46.8	802	US-10-104-774-2	Sequence 2, Appli

13	35.6	46.8	802	16	US-10-455-150-2	Sequence 2, Appli
14	35.6	46.8	852	18	US-10-723-860-5797	Sequence 5797, Ap
C 15	28.8	37.9	129	17	US-10-437-963-53858	Sequence 53858, A
C 16	28.2	37.1	225883	14	US-10-175-523-57	Sequence 57, Appl
C 17	27.6	36.3	1183	16	US-10-432-241-6	Sequence 6, Appli
C 18	27.6	36.3	1881	16	US-10-432-241-2	Sequence 2, Appli
C 19	26.8	35.3	849	16	US-10-389-566-110	Sequence 150, App
C 20	26.6	35.0	198161	17	US-10-775-169-52	Sequence 52, Appl
C 21	26.6	35.0	198161	18	US-10-723-860-165	Sequence 165, App
C 22	26.2	34.5	640	17	US-10-767-701-3691	Sequence 3691, Ap
C 23	26	34.2	795	16	US-10-264-237-748	Sequence 748, App
C 24	25.8	33.9	2415	18	US-10-425-115-114894	Sequence 114894, A
C 25	25.6	33.7	654	13	US-10-027-632-33959	Sequence 33959, A
C 26	25.6	33.7	654	15	US-10-027-632-33959	Sequence 33959, A
C 27	25.6	33.7	739	13	US-10-027-632-146671	Sequence 146671, A
C 28	25.6	33.7	739	13	US-10-027-632-146672	Sequence 146672, A
C 29	25.6	33.7	739	15	US-10-027-632-146671	Sequence 146671, A
C 30	25.6	33.7	739	15	US-10-027-632-146672	Sequence 146672, A
C 31	25.6	33.7	763	13	US-10-027-632-170284	Sequence 170284, A
C 32	25.6	33.7	763	15	US-10-027-632-170284	Sequence 170284, A
C 33	25.4	33.4	818	13	US-10-027-632-9912	Sequence 9912, Ap
C 34	25.4	33.4	818	13	US-10-027-632-9913	Sequence 9913, Ap
C 35	25.4	33.4	818	15	US-10-027-632-9912	Sequence 9912, Ap
C 36	25.4	33.4	818	15	US-10-027-632-9913	Sequence 9913, Ap
C 37	25.4	33.4	2025	17	US-10-437-963-73382	Sequence 73382, A
C 38	25.2	33.2	183	18	US-10-674-124A-4652	Sequence 4652, Ap
C 39	25.2	33.2	1339	16	US-10-425-114-31894	Sequence 31894, A
C 40	25.2	33.2	2154	18	US-10-425-115-37543	Sequence 37543, A
C 41	24.8	32.6	540	13	US-10-027-632-307914	Sequence 307914, A
C 42	24.8	32.6	540	15	US-10-027-632-307914	Sequence 307914, A
C 43	24.6	32.4	405	14	US-10-198-846-3322	Sequence 3322, Ap
C 44	24.6	32.4	488	17	US-10-437-963-56587	Sequence 56587, A
C 45	24.6	32.4	676	13	US-10-027-632-237425	Sequence 237425, A

#### ALIGNMENTS

RESULT 1  
US-10-027-632-287049/c  
; Sequence 287049, Application US/10027632  
; Publication No. US20020198371A1  
GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE REFERENCE: Polymorphisms in the Human Genome  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 287049  
; LENGTH: 600  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-287049  
Query Match 68.4%; Score 52; DB 13; Length 600;  
Best Local Similarity 91.7%; Pred. No. 2.5e-10;

Matches 55; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 GATTTGTTGCTGCTCTCTGGAGAGAGCTTTGCTATGCCCGTAGTAAATACCCCT 60  
|||||  
Db 245 GATTTATTTGCTGCTCTCTGGAGAGAGCTTTGCTATGCCCGTAGTAAATACCCCT 186

RESULT 2  
US-10-027-632-287050/c

; Sequence 287050, Application US/10027632  
; Publication No. US20020198371A1

; GENERAL INFORMATION:

; APPLICANT: Wang, David G.

; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide

; FILE REFERENCE: 108827.129

; CURRENT APPLICATION NUMBER: US/10/027,632

; PRIOR FILING DATE: 2002-04-30

; PRIOR APPLICATION NUMBER: US 60/218,006

; PRIOR FILING DATE: 2000-07-12

; PRIOR APPLICATION NUMBER: US 60/198,676

; PRIOR FILING DATE: 2000-04-20

; PRIOR APPLICATION NUMBER: US 60/193,483

; PRIOR FILING DATE: 2000-03-29

; PRIOR APPLICATION NUMBER: US 60/185,218

; PRIOR FILING DATE: 2000-02-24

; PRIOR APPLICATION NUMBER: US 60/167,363

; PRIOR FILING DATE: 1999-11-23

; PRIOR APPLICATION NUMBER: US 60/156,358

; PRIOR FILING DATE: 1999-09-28

; PRIOR APPLICATION NUMBER: US 60/146,002

; PRIOR FILING DATE: 1999-08-09

; NUMBER OF SEQ ID NOS: 325720

; SOFTWARE: FASTSEQ for Windows Version 4.0

; SEQ ID NO 287050

; LENGTH: 600

; TYPE: DNA

; ORGANISM: Human

US-10-027-632-287050

Query Match 68.4%; Score 52; DB 13; Length 600;

Best Local Similarity 91.7%; Pred. No. 2.5e-10;

Matches 55; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 GATTTGTTGCTGCTCTCTGGAGAGAGCTTTGCTATGCCCGTAGTAAATACCCCT 60  
|||||  
Db 245 GATTTATTTGCTGCTCTCTGGAGAGAGCTTTGCTATGCCCGTAGTAAATACCCCT 186

RESULT 3

US-10-027-632-287049/c

; Sequence 287049, Application US/10027632

; Publication No. US20030204075A9

; GENERAL INFORMATION:

; APPLICANT: Wang, David G.

; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide

; FILE REFERENCE: 108827.129

; CURRENT APPLICATION NUMBER: US/10/027,632

; PRIOR FILING DATE: 2002-04-30

; PRIOR APPLICATION NUMBER: US 60/218,006

; PRIOR FILING DATE: 2000-07-12

; PRIOR APPLICATION NUMBER: US 60/198,676

; PRIOR FILING DATE: 2000-04-20

; PRIOR APPLICATION NUMBER: US 60/193,483

; PRIOR FILING DATE: 2000-03-29

; PRIOR APPLICATION NUMBER: US 60/185,218

; PRIOR FILING DATE: 2000-02-24

; PRIOR APPLICATION NUMBER: US 60/167,363

; PRIOR FILING DATE: 1999-11-23

; PRIOR APPLICATION NUMBER: US 60/156,358

; PRIOR FILING DATE: 1999-09-28

; PRIOR APPLICATION NUMBER: US 60/146,002

; PRIOR FILING DATE: 1999-08-09

; NUMBER OF SEQ ID NOS: 325720

; SOFTWARE: FASTSEQ for Windows Version 4.0

; SEQ ID NO 287049

; LENGTH: 600

; TYPE: DNA

; ORGANISM: Human

US-10-027-632-287049

Query Match 68.4%; Score 52; DB 15; Length 600;

Best Local Similarity 91.7%; Pred. No. 2.5e-10;

Matches 55; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 GATTTGTTGCTGCTCTCTGGAGAGAGCTTTGCTATGCCCGTAGTAAATACCCCT 60  
|||||  
Db 245 GATTTATTTGCTGCTCTCTGGAGAGAGCTTTGCTATGCCCGTAGTAAATACCCCT 186

RESULT 4  
US-10-027-632-287050/c

; Sequence 287050, Application US/10027632

; Publication No. US20030204075A9

; GENERAL INFORMATION:

; APPLICANT: Wang, David G.

; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide

; FILE REFERENCE: 108827.129

; CURRENT APPLICATION NUMBER: US/10/027,632

; PRIOR FILING DATE: 2002-04-30

; PRIOR APPLICATION NUMBER: US 60/218,006

; PRIOR FILING DATE: 2000-07-12

; PRIOR APPLICATION NUMBER: US 60/198,676

; PRIOR FILING DATE: 2000-04-20

; PRIOR APPLICATION NUMBER: US 60/193,483

; PRIOR FILING DATE: 2000-03-29

; PRIOR APPLICATION NUMBER: US 60/185,218

; PRIOR FILING DATE: 2000-02-24

; PRIOR APPLICATION NUMBER: US 60/167,363

; PRIOR FILING DATE: 1999-11-23

; PRIOR APPLICATION NUMBER: US 60/156,358

; PRIOR FILING DATE: 1999-09-28

; PRIOR APPLICATION NUMBER: US 60/146,002

; PRIOR FILING DATE: 1999-08-09

; NUMBER OF SEQ ID NOS: 325720

; SOFTWARE: FASTSEQ for Windows Version 4.0

; SEQ ID NO 287050

; LENGTH: 600

; TYPE: DNA

; ORGANISM: Human

US-10-027-632-287050

Query Match 68.4%; Score 52; DB 15; Length 600;

Best Local Similarity 91.7%; Pred. No. 2.5e-10;

Matches 55; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 GATTTGTTGCTGCTCTCTGGAGAGAGCTTTGCTATGCCCGTAGTAAATACCCCT 60  
|||||  
Db 245 GATTTATTTGCTGCTCTCTGGAGAGAGCTTTGCTATGCCCGTAGTAAATACCCCT 186

RESULT 5  
US-10-723-860-1311

; Sequence 1311, Application US/10723860

; Publication No. US20040253606A1

; GENERAL INFORMATION:

; APPLICANT: Aziz, Natasha

; APPLICANT: Ginsburg, Wendy M.

; APPLICANT: Zlotnik, Albert

; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &

; FILE REFERENCE: 05882.0193.NPUS01

; CURRENT APPLICATION NUMBER: US/10/723,860

; PRIOR FILING DATE: 2003-11-26





Query Match	48.9%; Score 37.2; DB 16; Length 793;
-------------	---------------------------------------

RESULT 13  
US-10-455-150-2  
; Sequence 2, Application US/10455150

```
/ Publication No. US20040053302A1
/ GENERAL INFORMATION:
/ APPLICANT: Perkin-Elmer Corporation,
/ TITLE OF INVENTION: Applied Biosystems Division
/ SITES
/ NUMBER OF SEQUENCES: 11
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: David J. Weltz,
/ Wilson Sonsini Goodrich & Rosati
/ STREET: 650 Page Mill Road
/ CITY: Palo Alto
/ STATE: California
/ COUNTRY: USA
/ ZIP: 94304-1050
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: 3.5 inch diskette
/ COMPUTER: IBM compatible
/ OPERATING SYSTEM: Microsoft Windows 95
/ SOFTWARE: Wordperfect for windows 6.0,
/ ASCII (DOS) TEXT format
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/10/455,150
/ FILING DATE: 04-Jun-2003
/ CLASSIFICATION: <Unknown>
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US/09/324,709A
/ FILING DATE: 03-June-1999
/ ATTORNEY/AGENT INFORMATION:
/ NAME: David J. Weltz
/ REGISTRATION NUMBER: 38,362
/ REFERENCE/DOCKET NUMBER: 16842-758
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (650) 493-9300
/ TELEFAX: (650) 493-6811
/ INFORMATION FOR SEQ ID NO: 2:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 802 nucleotides
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-10-455-150-2

Query Match          46.8%; Score 35.6; DB 16; Length 802;
Best Local Similarity 90.5%; Pred. No. 0.001;
Matches 38; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1 GATTTGTTGCTGCTGCTGGAGAGCCTTGTGCTATGCC 42
DB      80 GATTTGTTGCTGCTGCTGGAGAGCCTTGTGCTATGCC 121

RESULT 14
US-10-723-860-5797
/ Sequence 5797, Application US/10723860
/ Publication No. US20040253606A1
/ GENERAL INFORMATION:
/ APPLICANT: Aziz, Natasha
/ APPLICANT: Gansburg, Wendy M.
/ APPLICANT: Zlotnick, Albert
/ TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
/ FILE REFERENCE: 05882.0193.NPUS01
/ CURRENT APPLICATION NUMBER: US/10/723,860
/ PRIOR FILING DATE: 2003-11-26
/ PRIOR APPLICATION NUMBER: 60/429,739
/ PRIOR FILING DATE: 2002-11-26
/ NUMBER OF SEQ ID NOS: 8393
/ SOFTWARE: PatentIn version 3.2
/ SEQ ID NO 5797
/ LENGTH: 852
/ TYPE: DNA
```

```
/ ORGANISM: Homo sapiens
US-10-723-860-5797

Query Match          46.8%; Score 35.6; DB 18; Length 852;
Best Local Similarity 90.5%; Pred. No. 0.001;
Matches 38; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1 GATTTGTTGCTGCTGCTGGAGAGCCTTGTGCTATGCC 42
DB      80 GATTTGTTGCTGCTGCTGGAGAGCCTTGTGCTATGCC 121

RESULT 15
US-10-437-963-53858/c
/ Sequence 53858, Application US/10437963
/ Publication No. US20040123343A1
/ GENERAL INFORMATION:
/ APPLICANT: La Rosa, Thomas J.
/ APPLICANT: Kovacic, David K.
/ APPLICANT: Zhou, Yihua
/ APPLICANT: Cao, Yongwei
/ APPLICANT: Wu, Wei
/ APPLICANT: Boukharov, Andrey A.
/ APPLICANT: Barbazuk, Brad
/ APPLICANT: Li, Ping
/ TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
/ FILE REFERENCE: 38-21(53221)B
/ CURRENT APPLICATION NUMBER: US/10/437,963
/ CURRENT FILING DATE: 2003-05-14
/ NUMBER OF SEQ ID NOS: 204966
/ SEQ ID NO 53858
/ LENGTH: 129
/ TYPE: DNA
/ ORGANISM: Oryza sativa
/ FEATURE:
/ OTHER INFORMATION: Clone ID: PAT_MRT4530_56017C.1
US-10-437-963-53858

Query Match          37.9%; Score 28.8; DB 17; Length 129;
Best Local Similarity 69.6%; Pred. No. 0.3;
Matches 39; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY      3 TTTTGTGCTGCTGCTGCTGGAGAGCCTTGTGCTATGCCGTTGATTAATATGCC 58
DB      118 TATTTGCTGCTGCTGCTGGAGATGAGTTCTTCATGCTCAATTAAGAACCC 63

Search completed: February 9, 2005, 06:50:58
Job time : 104.41 secs
```

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 8, 2005, 12:58:25 ; Search time 905.232 Seconds  
(without alignments)  
3059.349 Million cell updates/sec

Title: US-10-754-437-6

Perfect score: 1 gattctgttcgtcctgcctcc.....ccctgcataataatcattc 76

Sequence: 1 gattctgttcgtcctgcctcc.....ccctgcataataatcattc 76

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 32822875 seqs, 18219665908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Database : Maximum Match 100%

EST: \*  
1: gb\_ests1:\*  
2: gb\_ests2:\*  
3: gb\_ests3:\*  
4: gb\_ests4:\*  
5: gb\_ests5:\*  
6: gb\_ests6:\*  
7: gb\_ests7:\*  
8: gb\_ests8:\*  
9: gb\_ests9:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	41.8	55.0	303	W33932	mb54d02.r1
2	41.8	55.0	430	W40649	mc43h12.r1
3	41.8	55.0	536	W36345	mb72c12.r1
4	41.8	55.0	585	AY419441	Mus muscu
5	41.8	55.0	645	CD772419	AGENCOURT
6	41.8	55.0	707	CB056709	NISC_j119
7	41.8	55.0	812	AK029358	Mus muscu
8	41.8	55.0	843	CB588212	AGENCOURT
9	41.8	55.0	859	CB588525	AGENCOURT
10	41.8	55.0	865	CB589177	AGENCOURT
11	41.8	55.0	886	CB587051	AGENCOURT
12	41.8	55.0	889	CB589251	AGENCOURT
13	41.8	55.0	891	CB574837	AGENCOURT
14	41.8	55.0	913	CB590451	AGENCOURT
15	41.8	55.0	919	CB587332	AGENCOURT
16	41.8	55.0	928	CB590111	AGENCOURT
17	41.8	55.0	945	BB614068	BB614068
18	41.8	55.0	990	W12906	ma89b03.r1
19	41.8	55.0	1006	W29475	mb99f11.r1
20	41.8	55.0	1020	W08102	mb40g08.r1
21	40.2	52.9	217	R46913	R46913
22	40.2	52.9	223	BB571643	BB571643
23	40.2	52.9	449	R47024	R47024
24	40.2	52.9	1031	CB575508	CB575508

25	39.2	51.6	373	7	R47135	R47135
26	39.2	51.6	395	7	R47030	R47030
27	39.2	51.6	556	7	R46934	R46934
28	38.6	50.8	257	7	R47100	R47100
29	38.2	50.3	303	7	R46903	R46903
30	38.2	50.3	318	7	R47078	R47078
31	38.2	50.3	467	7	R46947	R46947
32	37.6	49.5	293	7	R47143	R47143
33	37.2	48.9	429	6	CB473334	CB473334
34	37.2	48.9	570	9	AY419439	AY419439
35	37.2	48.9	570	9	AY419440	AY419440
36	37	48.7	395	7	R46955	R46955
37	36.6	48.2	478	7	R46933	R46933
38	36.2	47.6	273	7	R47002	R47002
39	35.6	46.8	286	7	R46887	R46887
40	35.6	46.8	783	4	BG188114	BG188114
41	35	46.1	355	7	R46940	R46940
42	35	46.1	355	7	R46941	R46941
43	33.6	44.2	748	8	CC324394	CC324394
44	33.6	44.2	1180	8	CC301510	CC301510
45	33	43.4	295	2	BB61682	BB61682

## ALIGNMENTS

RESULT 1  
W33932  
LOCUS  
DEFINITION  
mb54d02.r1 Soares mouse p3NMF19.5 Mus musculus cDNA clone  
IMAGE:333219 5' similar to gb:M10095 Mouse amelogenin (MOUSE);  
RNA sequence.  
W33932  
W33932.1 GI:1315837  
EST.  
Mus musculus (house mouse)  
Mus musculus  
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 303)  
Marras, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
Waterston, R.  
The WashU-HMI Mouse EST Project  
Unpublished (1996)  
Contact: Maria M/Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LIND, contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
WGI:214619  
Seq primer: mob.RBGA\*ET  
High quality sequence stop: 185.  
Location/Qualifiers  
1..303  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:333219"  
/dev\_stage="19.5 dpc total fetus"  
/lab\_host="DH10B (ampicillin resistant)"  
/clone\_1lb="Soares mouse p3NMF19.5"  
/note="Vector: pRT73D (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Bco RI; 1st strand cDNA  
was primed with a Not I - oligo(dT) primer [5'  
TGTATACCAATCTGAAGTGGAGCGCCGCAATTTTATTTTATTTT 3']  
double-stranded cDNA was size selected, ligated to Bco RI

adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M. Patricia Bonaldo. RNA was kindly provided by Dr. Minoru Ko (Wayne State University)."

## ORIGIN

Query Match 55.0%; Score 41.8; DB 7; Length 303;  
Best Local Similarity 95.6%; Pred. No. 0.00046;  
Matches 43; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GATTTGTTGGCTGCTGCTGCTGAGAGAGCTTTGCTATGCCCGT 45  
|||||  
Db 81 GATTTGTTGGCTGCTGCTGAGAGAGCTTTGCTATGCCCGT 125

## RESULT 2

W40649 430 bp mRNA linear EST 11-SEP-1996  
LOCUS me43h12.r1 Soares mouse p3NMF19.5 Mus musculus cDNA clone  
DEFINITION IMAGE:351335 5' similar to gb:M10095 Mouse amelogenin (MOUSE);,  
mRNA sequence.

ACCESSION W40649 GI:1324982  
VERSION W40649.1  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 430)  
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisels,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Scheißenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.

TITLE The WashU-HMNI Mouse EST Project  
JOURNAL Unpublished (1996)  
COMMENT Contact: Marra M/Mouse EST Project  
WashU-HMNI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:223135  
Seq primer: ETPprimer  
High quality sequence stop: 323.

## FEATURES

SOURCE  
Location/Qualifiers

1..430  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:351335"  
/dev\_stage="19.5 dpc total fetus"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Vector: pT73D (Pharmacia) with a modified polylinker. Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTGAAGTGGAGGCGCGCCATTTTCTTTTCTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M. Patricia Bonaldo. RNA was kindly provided by Dr. Minoru Ko (Wayne State University)."

## ORIGIN

Query Match

55.0%; Score 41.8; DB 7; Length 430;

Best Local Similarity 95.6%; Pred. No. 0.0005;  
Matches 43; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GATTTGTTGGCTGCTGCTGAGAGAGCTTTGCTATGCCCGT 45  
|||||  
Db 69 GATTTGTTGGCTGCTGCTGAGAGAGCTTTGCTATGCCCGT 113

## RESULT 3

W36345 536 bp mRNA linear EST 11-SEP-1996  
LOCUS mb72c12.r1 Soares mouse p3NMF19.5 Mus musculus cDNA clone  
DEFINITION IMAGE:334966 5' similar to gb:M10095 Mouse amelogenin (MOUSE);,  
mRNA sequence.

ACCESSION W36345 GI:1318120  
VERSION W36345.1  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 536)  
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisels,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Scheißenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.

TITLE The WashU-HMNI Mouse EST Project  
JOURNAL Unpublished (1996)  
COMMENT Contact: Marra M/Mouse EST Project  
WashU-HMNI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:216366  
Seq primer: ETPprimer  
High quality sequence stop: 359.

## FEATURES

SOURCE  
Location/Qualifiers

1..536  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:334966"  
/dev\_stage="19.5 dpc total fetus"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Vector: pT73D (Pharmacia) with a modified polylinker. Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTGAAGTGGAGGCGCGCCATTTTCTTTTCTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M. Patricia Bonaldo. RNA was kindly provided by Dr. Minoru Ko (Wayne State University)."

## ORIGIN

Query Match 55.0%; Score 41.8; DB 7; Length 536;  
Best Local Similarity 95.6%; Pred. No. 0.00053;  
Matches 43; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GATTTGTTGGCTGCTGCTGAGAGAGCTTTGCTATGCCCGT 45  
|||||  
Db 64 GATTTGTTGGCTGCTGCTGAGAGAGCTTTGCTATGCCCGT 108

## RESULT 4

LOCUS	AY19441	585 bp	DNA	linear	GSS 17-DEC-2003
DEFINITION	Mus musculus AMELX gene, VIRTUAL TRANSCRIPT, partial sequence, genomic survey sequence.				
ACCESSION	AY19441				
VERSION	AY19441.1	GI:39775398			
KEYWORDS	GSS.				
SOURCE	Mus musculus (house mouse)				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
REFERENCE	1 (bases 1 to 585)				
AUTHORS	Clark,A.G., Gianowski,S., Nielson,R., Thomas,P., Kejarival,A., Todd,M.A., Tenenbaum,D.M., Civeello,D.R., Lu,F., Murphy,B., Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J., Adams,M.D. and Cargill,M.				
TITLE	Interfering nonneutral evolution from human-chimp-mouse orthologous gene trios				
JOURNAL	Science 302 (5652), 1960-1963 (2003)				
PUBMED	14671302				
REFERENCE	2 (bases 1 to 585)				
AUTHORS	Clark,A.G., Gianowski,S., Nielson,R., Thomas,P., Kejarival,A., Todd,M.A., Tenenbaum,D.M., Civeello,D.R., Lu,F., Murphy,B., Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J., Adams,M.D. and Cargill,M.				
TITLE	Direct Submission				
JOURNAL	Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA				
COMMENT	These sequences were made by sequencing genomic exons and ordering them based on alignment.				
FEATURES	Location/Qualifiers				
SOURCE	1..585				
gene	/organism="Mus musculus"				
	/mol_type="Genomic DNA"				
	/db_xref="taxon:10090"				
	<1..->585				
	/gene="AMELX"				
	/locus_tag="HCK6891"				
ORIGIN					
Query Match	55.0%; Score 41.8; DB 9; Length 585;				
Beat Local Similarity	95.6%; Pred. NO. 0.00054;				
Matches	43; Conservative 0; Mismatches 2; Indels 0; Gaps 0;				
Qy	1 GATTTGTTGGCTGCTCTCGGAGCAGCCCTTGCTATGCCCT 45				
Db	12 GATTTGTTGGCTGCTCTCGGAGCAGCCCTTGCTATGCCCT 56				
RESULT 5					
LOCUS	CD773419	645 bp	mRNA	linear	EST 02-JUL-2003
DEFINITION	AGNCOURT 14713373 NIH MGC 190 Mus musculus cDNA clone IMAGE:30501267 5', mRNA sequence.				
ACCESSION	CD773419				
VERSION	CD773419.1	GI:32431921			
KEYWORDS	EST.				
SOURCE	Mus musculus (house mouse)				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
REFERENCE	1 (bases 1 to 645)				
AUTHORS	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .				
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)				
JOURNAL	Unpublished (1999)				
COMMENT	Contact: Daniele S. Gethard, Ph.D. Office of Cancer Genomics National Cancer Institute / NIH Bldg. 31 Rm10A07 Bethesda, MD 20892 Email: <a href="mailto:cgapdb-remail.nih.gov">cgapdb-remail.nih.gov</a> Tissue Procurement: Yoshihiko Yamada, Takashi Nakamura, NIDCR cDNA Library Preparation: CLONTECH Laboratories, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)				

DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.llnl.gov>  
 Plate: NDCM223 row: k column: 04  
 High quality sequence stop: 611.  
 Location/Qualifiers  
 1..645  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:30501267"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_id="NH\_MGC\_190"  
 /note="Organ: Pooled - Molar; Vector: pDNR-LIB; Site: 1:  
 SfiI (ggcgccctggcc); Site 2: SfiI (ggcgccctggcc);  
 Non-normalized full-length enriched library 5' and 3',  
 adaptors were used in cloning as follows: 5' adaptor  
 sequence: 5'-CACGCGCATTTATGACC-3' and 3' adaptor sequence:  
 5'-ATTCTAGAGCGCCGAGCGCCGACATG-dt(30)BN-3' (where B = A,  
 C, or G and N = A, C, G, or T). Average insert size 1.71  
 kb (range 0.5-3.0 kb). 15/15 colonies contained inserts by  
 PCR. This library was enriched for full-length clones and  
 was constructed by Clontech Laboratories (Palo Alto, CA)  
 Corp."

ORIGIN  
 Query Match 55.0%; Score 41.8; DB 6; Length 645;  
 Best Local Similarity 95.6%; Pred. No. 0.00056;  
 Matches 43; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GATTTCGTTGCTGCTGCTCTCGAGAGACCTTGTATGCCCCGT 45  
 |||||  
 Db 84 GATTTCGTTGCTGCTGCTCTCGAGAGACCTTGTATGCCCCCT 128  
 |||||

RESULT 6 707 bp mRNA linear EST 17-JAN-2003  
 LOCUS CB056709  
 DEFINITION NISC j119g05.w1 Soares NMNP13-15 Mus musculus cDNA clone  
 IMAGE:4848584 5', mRNA sequence.  
 CB056709  
 CB056709.1 GI:27794996  
 EST.  
 Mus musculus (house mouse)  
 Mus musculus  
 Eukaryotes; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 707)  
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished (1997)  
 Contact: Robert Strauberg, Ph.D.  
 Email: [cgapds-remail.nih.gov](mailto:cgapds-remail.nih.gov)  
 CDNA Library Preparation:  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium/LNL  
 DNA Sequencing by: National Institutes of Health Intramural  
 Sequencing Center (NISC)  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.llnl.gov>  
 Plate: LHAM10779 row: N column: 9  
 Seq primer: T7 primer.  
 Location/Qualifiers  
 1..707  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:4848584"  
 /tissue\_type="pituitary gland"  
 /dev\_stage="juvenile, 13-15 days"  
 /lab\_host="DH10B (phage-resistant)"

FEATURES  
 SOURCE





SOURCE	Mus musculus (house mouse)
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclirognathi; Muridae; Murinae; Mus.
REFERENCE	1 (bases 1 to 843)
AUTHORS	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished (1999)
COMMENT	Contact: Robert Strausberg, Ph.D. Email: <a href="mailto:cgapbs-remail.nih.gov">cgapbs-remail.nih.gov</a> Tissue Procurement: Dr. David Rowe cDNA Library Preparation: Invitrogen Corp DNA Sequencing by: Agencourt Bioscience Corporation (LNL) Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a> Plate: NDAM338 row: g column: 08 High quality sequence crop: 636. Location/Qualifiers
FEATURES	1..843 <code>organism="Mus musculus"</code> <code>mol_type="mRNA"</code> <code>db_xref="taxon:10090"</code> <code>clone="IMAGE:30293239"</code> <code>tissue_type="embryonic limb, maxilla and mandible"</code> <code>lab_host="DH10B (phage-resistant)"</code> <code>clone_1lb="NIH MGC 136"</code> <code>note="Vector: pCMV-SPORT6.1; Site 1: EcoRV; Site 2: NotI</code> Normalized, full-length enriched library from pool of mouse embryonic limb, maxilla and mandible, embryonic day 17.5, 18.5 and newborn (mandible (5), 4 and 1 limb and jaw equivalents from respective days). Cloned directionally, oligo-dT primed (5'-GACTAGTTCTGATCGCAGCGCCGCC (7)15-3', size selected for the >1kb fragments, average insert size 1.2 kb. Normalization to Cot 7.5. Tissue contributed by David Rowe; library constructed by Reegen, Invitrogen Corp. Note: this is a NIH_MGC library."
ORIGIN	
Query Match	55.0%; Score 41.8; DB 6; Length 843;
Best Local Similarity	95.6%; Pred No. 0.0066;
Matches	43; Conservative 0; Mismatches 2; Indels 0; Gaps 0
Dn	1 GATTTCCTTCCTCCTCTGAGAGACAGCTTTCCTAGCCCGT 45 51 GATTTCCTTCCTCCTCCTGAGAGACAGCTTTCCTAGCCCGT 95
RESULT 9	
CB588525	859 bp mRNA linear EST 03-APR-2003
LOCUS	AGENCOURT_12567944 NIH MGC_136 Mus musculus cDNA clone
DEFINITION	IMAGE:30289751 5', mRNA sequence.
ACCESSION	CB588525
VERSION	CB588525.1 GI:29506381
KEYWORDS	EST.
SOURCE	Mus musculus (house mouse)
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclirognathi; Muridae; Murinae; Mus.
REFERENCE	1 (bases 1 to 859)
AUTHORS	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished (1999)
COMMENT	Contact: Robert Strausberg, Ph.D. Email: <a href="mailto:cgapbs-remail.nih.gov">cgapbs-remail.nih.gov</a> Tissue Procurement: Dr. David Rowe cDNA Library Preparation: Invitrogen Corp DNA Sequencing by: Agencourt Bioscience Corporation (LNL) Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:

```

http://image.lnl.gov
plate: NDAM329 row: e column: 24
High quality sequence stop: 763.
Location/Qualifiers
1..859
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:30289751"
/tissue_type="embryonic limb, maxilla and mandible"
/lab_host="DHI0B (phage-resistant)"
/clone_lib="NIH MGC 136"
/notes="Vector: pCMV-Sport6.1; Site 1: EcoRV; Site 2: NotI; Normalized, full-length enriched library from pool of mouse embryonic limb, maxilla and mandible, embryonic day 17.5, 18.5 and newborn (mandible (5, 4 and 1 limb and jaw equivalents from respective days). Cloned directionally, oligo-dT primed (5'-GACAGTCTGATGCGGAGCGCGCCCT)15-3'. Size selected for the >1kb fragments, average insert size 1.2 kb. Normalization to Cot 7.5. Tissue contributed by David Rowe; library constructed by ResGen, Invitrogen Corp. Note: this is a NIH_MGC Library."

ORIGIN

Query Match 55.0%; Score 41.8; DB 6; Length 859;
Best Local Similarity 95.6%; Pred. No. 0.0006;
Matches 43; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 GATTTGTTGCGTCGCGCTCGGAGAGAGCGCTTGATGCCCGT 45
|||||
70 GATTTGTTGCGTCGCTCGGAGAGAGCGCTTGATGCCCGCT 114

RESULT 10
CB589177
LOCUS CB589177
DEFINITION AGENCOURT 12770446 NIH_MGC_136 Mus musculus cDNA clone
ACCESSION CB589177
VERSION IMAG:30290534 5', mRNA sequence.
KEYWORDS CB589177.1 GI:29507033
SOURCE BSR.
ORGANISM Mus musculus (house mouse)
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 865)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Dr. David Rowe
CDNA Library Preparation: Invitrogen Corp
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: NDAM331 row: f column: 15
High quality sequence stop: 630.
Location/Qualifiers
1..865
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:30290534"
/tissue_type="embryonic limb, maxilla and mandible"
/lab_host="DHI0B (phage-resistant)"
/clone_lib="NIH MGC 136"
/notes="Vector: pCMV-Sport6.1; Site 1: EcoRV; Site 2: NotI; Normalized, full-length enriched library from pool of mouse embryonic limb, maxilla and mandible, embryonic day 17.5, 18.5 and newborn (mandible (5, 4 and 1 limb and jaw

```

equivalents from respective days). Cloned directionally, oligo-dT primed (5'-GACTAGTCTAGATCGGAGCGGCCCTT)15-3'. Size selected for the 1kb fragments, average insert size 1.2 kb. Normalization to Cot 7.5. Tissue contributed by David Rowe; library constructed by ResGen, Invitrogen Corp. Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 55.0%; Score 41.8; DB 6; Length 865;  
Best Local Similarity 95.6%; Pred. No. 0.0006;  
Matches 43; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GATTGTGTTGCTGCTCTCTGAGAGAGCCCTTGTATGCCCT 45  
DB 69 GATTGTGTTGCTGCTCTCTGAGAGAGCCCTTGTATGCCCT 113

RESULT 11  
LOCUS CB587051  
DEFINITION AGENCOURT 12807074 NIH MGC 136 Mus musculus CDNA clone  
IMAGE:30295123 5', mRNA sequence.  
ACCESSION CB587051  
VERSION CB587051.1 GI:29504907  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 886)  
REFERENCE NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgaabs-remail.nih.gov  
Tissue Procurement: Dr. David Rowe  
CDNA Library Preparation: Invitrogen Corp  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Plate: NDAM343 row: e column: 20  
High quality sequence stop: 619.  
Location/Qualifiers  
1. 886  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:30295123"  
/issue\_type="embryonic limb, maxilla and mandible"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH MGC 136"  
/note="Vector: PCMV-SPORT6.1; Site 1: EcoRV; Site 2: NotI; Normalized, full-length enriched library from pool of mouse embryonic limb, maxilla and mandible, embryonic day 17.5, 18.5 and newborn (mandible (5, 4 and 1 limb and jaw equivalents from respective days). Cloned directionally, oligo-dT primed (5'-GACTAGTCTAGATCGGAGCGGCCCTT)15-3'. Size selected for the >1kb fragments, average insert size 1.2 kb. Normalization to Cot 7.5. Tissue contributed by David Rowe; library constructed by ResGen, Invitrogen Corp. Note: this is a NIH\_MGC Library."

FEATURES  
source

1. 886  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:30295123"  
/issue\_type="embryonic limb, maxilla and mandible"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH MGC 136"  
/note="Vector: PCMV-SPORT6.1; Site 1: EcoRV; Site 2: NotI; Normalized, full-length enriched library from pool of mouse embryonic limb, maxilla and mandible, embryonic day 17.5, 18.5 and newborn (mandible (5, 4 and 1 limb and jaw equivalents from respective days). Cloned directionally, oligo-dT primed (5'-GACTAGTCTAGATCGGAGCGGCCCTT)15-3'. Size selected for the >1kb fragments, average insert size 1.2 kb. Normalization to Cot 7.5. Tissue contributed by David Rowe; library constructed by ResGen, Invitrogen Corp. Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 55.0%; Score 41.8; DB 6; Length 865;  
Best Local Similarity 95.6%; Pred. No. 0.0006;  
Matches 43; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GATTGTGTTGCTGCTCTCTGAGAGAGCCCTTGTATGCCCT 45  
DB 95 GATTGTGTTGCTGCTCTCTGAGAGAGCCCTTGTATGCCCT 139

RESULT 12  
LOCUS CB589251  
DEFINITION AGENCOURT 12780700 NIH MGC 136 Mus musculus CDNA clone  
IMAGE:30287069 5', mRNA sequence.

ACCESSION CB589251  
VERSION CB589251.1 GI:29507107  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 889)  
REFERENCE NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgaabs-remail.nih.gov  
Tissue Procurement: Dr. David Rowe  
CDNA Library Preparation: Invitrogen Corp  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Plate: NDAM322 row: f column: 06  
High quality sequence start: 23  
High quality sequence stop: 581.  
Location/Qualifiers  
1. 889  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:30287069"  
/issue\_type="embryonic limb, maxilla and mandible"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH MGC 136"  
/note="Vector: PCMV-SPORT6.1; Site 1: EcoRV; Site 2: NotI; Normalized, full-length enriched library from pool of mouse embryonic limb, maxilla and mandible, embryonic day 17.5, 18.5 and newborn (mandible (5, 4 and 1 limb and jaw equivalents from respective days). Cloned directionally, oligo-dT primed (5'-GACTAGTCTAGATCGGAGCGGCCCTT)15-3'. Size selected for the >1kb fragments, average insert size 1.2 kb. Normalization to Cot 7.5. Tissue contributed by David Rowe; library constructed by ResGen, Invitrogen Corp. Note: this is a NIH\_MGC Library."

FEATURES  
source

1. 889  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:30287069"  
/issue\_type="embryonic limb, maxilla and mandible"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH MGC 136"  
/note="Vector: PCMV-SPORT6.1; Site 1: EcoRV; Site 2: NotI; Normalized, full-length enriched library from pool of mouse embryonic limb, maxilla and mandible, embryonic day 17.5, 18.5 and newborn (mandible (5, 4 and 1 limb and jaw equivalents from respective days). Cloned directionally, oligo-dT primed (5'-GACTAGTCTAGATCGGAGCGGCCCTT)15-3'. Size selected for the >1kb fragments, average insert size 1.2 kb. Normalization to Cot 7.5. Tissue contributed by David Rowe; library constructed by ResGen, Invitrogen Corp. Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 55.0%; Score 41.8; DB 6; Length 889;  
Best Local Similarity 95.6%; Pred. No. 0.0006;  
Matches 43; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GATTGTGTTGCTGCTCTCTGAGAGAGCCCTTGTATGCCCT 45  
DB 89 GATTGTGTTGCTGCTCTCTGAGAGAGCCCTTGTATGCCCT 133

RESULT 13  
LOCUS CB574837  
DEFINITION AGENCOURT 12973279 NIH MGC 136 Mus musculus CDNA clone  
IMAGE:30254719 5', mRNA sequence.  
ACCESSION CB574837  
VERSION CB574837.1 GI:29494367  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 891)  
REFERENCE NIH-MGC <http://mgc.nci.nih.gov/>.

QY 1 GATTGTGTTGCTGCTCTCTGAGAGAGCCCTTGTATGCCCT 45  
DB 95 GATTGTGTTGCTGCTCTCTGAGAGAGCCCTTGTATGCCCT 139

TITLE  
JOURNAL  
COMMENT  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-remail.nih.gov  
Tissue Procurement: Dr. David Rowe  
CDNA Library Preparation: Invitrogen Corp  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:  
http://image.llnl.gov  
Plate: NDAM342 row: d column: 24  
High quality sequence start: 19  
High quality sequence stop: 658.  
Location/Qualifiers  
1. .891  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:30294719"  
/tissue\_type="embryonic limb, maxilla and mandible"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_1lb="NIH\_MGC\_136"  
/note="Vector: PCMV-SPORT6.1; Site 1: EcoRV; Site 2: NotI; Normalized, full-length enriched library from pool of mouse embryonic limb, maxilla and mandible, embryonic day 17.5, 18.5 and newborn (mandible (5, 4 and 1 limb and jaw equivalents from respective days). Cloned directionally, oligo-dT primed (5'-GACTAGTCTAGATCGAGCGGCCGCTT)15-3'. Size selected for the >1kb fragments, average insert size 1.2 kb. Normalization to Cot 7.5. Tissue contributed by David Rowe; library constructed by Resgen, Invitrogen Corp. Note: this is a NIH\_MGC Library."

ORIGIN  
Query Match 55.0%; Score 41.8; DB 6; Length 891;  
Best Local Similarity 95.6%; Pred. No. 0.0006;  
Matches 43; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 GATTTGTTGCTGCTGCTGAGAGAGCGCTTGTATGCCCGT 45  
|||||  
90 GATTTGTTGCTGCTGCTGAGAGAGCGCTTGTATGCCCGT 134  
|||||

RESULT 14  
CBS90451 913 bp mRNA linear EST 03-APR-2003  
LOCUS AGENCOURT 12770290 NIH\_MGC\_136 Mus musculus CDNA clone  
IMAGE:30286944 5', mRNA sequence.  
CBS90451 GI:29508307  
EST.  
KEYWORDS  
SOURCE Mus musculus (house mouse)  
ORGANISM  
REFERENCE  
AUTHORS  
TITLES  
JOURNAL  
COMMENT  
Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-remail.nih.gov  
Tissue Procurement: Dr. David Rowe  
CDNA Library Preparation: Invitrogen Corp  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:  
http://image.llnl.gov  
Plate: NDAM322 row: a column: 01  
High quality sequence start: 12  
High quality sequence stop: 640.  
Location/Qualifiers

FEATURES  
source

source  
1. .913  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:30286944"  
/tissue\_type="embryonic limb, maxilla and mandible"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_1lb="NIH\_MGC\_136"  
/note="Vector: PCMV-SPORT6.1; Site 1: EcoRV; Site 2: NotI; Normalized, full-length enriched library from pool of mouse embryonic limb, maxilla and mandible, embryonic day 17.5, 18.5 and newborn (mandible (5, 4 and 1 limb and jaw equivalents from respective days). Cloned directionally, oligo-dT primed (5'-GACTAGTCTAGATCGAGCGGCCGCTT)15-3'. Size selected for the >1kb fragments, average insert size 1.2 kb. Normalization to Cot 7.5. Tissue contributed by David Rowe; library constructed by Resgen, Invitrogen Corp. Note: this is a NIH\_MGC Library."

ORIGIN  
Query Match 55.0%; Score 41.8; DB 6; Length 913;  
Best Local Similarity 95.6%; Pred. No. 0.00061;  
Matches 43; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 GATTTGTTGCTGCTGCTGAGAGAGCGCTTGTATGCCCGT 45  
|||||  
82 GATTTGTTGCTGCTGCTGAGAGAGCGCTTGTATGCCCGT 126  
|||||

RESULT 15  
CBS87332 919 bp mRNA linear EST 03-APR-2003  
LOCUS AGENCOURT 12971606 NIH\_MGC\_136 Mus musculus CDNA clone  
IMAGE:30291971 5', mRNA sequence.  
CBS87332 GI:29505188  
EST.  
KEYWORDS  
SOURCE Mus musculus (house mouse)  
ORGANISM  
REFERENCE  
AUTHORS  
TITLES  
JOURNAL  
COMMENT  
Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-remail.nih.gov  
Tissue Procurement: Dr. David Rowe  
CDNA Library Preparation: Invitrogen Corp  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:  
http://image.llnl.gov  
Plate: NDAM35 row: b column: 12  
High quality sequence start: 216  
High quality sequence stop: 325.  
Location/Qualifiers  
1. .919  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:30291971"  
/tissue\_type="embryonic limb, maxilla and mandible"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_1lb="NIH\_MGC\_136"  
/note="Vector: PCMV-SPORT6.1; Site 1: EcoRV; Site 2: NotI; Normalized, full-length enriched library from pool of mouse embryonic limb, maxilla and mandible, embryonic day 17.5, 18.5 and newborn (mandible (5, 4 and 1 limb and jaw equivalents from respective days). Cloned directionally, oligo-dT primed (5'-GACTAGTCTAGATCGAGCGGCCGCTT)15-3'. Size selected for the >1kb fragments, average insert size

FEATURES  
source



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OW nucleic - nucleic search, using sw model

Run on: February 8, 2005, 11:28:00 ; Search time 390.297 Seconds  
(without alignments)  
8481.439 Million cell updates/sec

Title: US-10-754-437-7

Perfect score: 70  
Sequence: 1 gattctgttcgctgcctcc.....aaattcccaattcccaattc 70

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_hcg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pac:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_srs:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	49.8	71.1	363	4	AB091789 Bos tauru
2	49.2	70.3	363	11	BV089294 RPAMSECO
3	49.2	70.3	363	11	BV097602 RPAMSECO
4	49.2	70.3	463	10	DB3063S2 Mus musculu
5	49.2	70.3	695	11	BV089295 RPAMSECO
6	49.2	70.3	695	11	BV097603 RPAMSECO
7	49.2	70.3	9384	10	AF294397 Mus muscu
8	49.2	70.3	95826	10	AL805974 Mouse DNA
9	47.6	68.0	212886	2	AC093946 Rattus no
10	47.6	68.0	234471	2	AC121424 Rattus no
11	46.8	66.9	8004	9	AB091786 Lemur cat
12	46	65.7	7425	4	AB091791 Sus scrofa
13	45	64.3	6442	9	AB091783 Saimiri b
14	43.6	62.3	5562	9	AB091787 Oculimur
15	43.6	62.3	5684	9	AB091785 Lemur cat
16	43.6	62.3	6465	9	AB091781 Pan trogl
17	43.6	62.3	8810	9	AY040206 Homo sapi
18	43.6	62.3	158142	2	AL357130 Homo sapi
19	43.6	62.3	259202	9	AC002366 Human Xp2

20	42.8	61.1	7163	9	AB091782 Pan trogl
21	42.8	61.1	38765	2	BS000568 Pan trogl
22	42.8	61.1	177654	2	AP000918 Homo sapi
23	42.8	61.1	177726	9	AC145770 Pan trogl
24	42.8	61.1	190089	9	BS000576 Pan trogl
25	42.8	61.1	200214	9	AC013412 Homo sapi
26	41.8	59.7	6264	4	AB091790 Bos tauru
27	40.8	58.3	5712	4	AB091793 Equus cab
28	40.2	57.4	746	4	AF215880 Capra hie
29	40.2	57.4	752	4	AF215889 Capra hie
30	40.2	57.4	757	4	BOVAMLGN1
31	40.2	57.4	820	4	BOVAMLGN1
32	38.6	55.1	778	4	SSU43405 Sus scrofa
33	38.6	55.1	6831	4	AB091792 Sus scrofa
34	37	52.9	7454	9	AB091784 Saimiri b
35	35.8	51.1	5591	4	AB091794 Equus cab
36	35.6	50.9	296	10	MUSMAMEB
37	35.6	50.9	727	10	MUSMAMEB
38	35.6	50.9	765	4	AB032194 Equus cab
39	35.6	50.9	789	4	AB032193 Equus cab
40	35.6	50.9	798	10	BC053090 Mus muscu
41	35.6	50.9	799	10	MUSMAMEA
42	35.6	50.9	825	6	AX865705 Sequence
43	35.6	50.9	825	6	BD145767 Primer fo
44	35.6	50.9	1626	6	AX882160 Sequence
45	35.6	50.9	1626	6	BD159567 Primer fo

## ALIGNMENTS

RESULT 1  
LOCUS AB091789 6451 bp DNA linear MAM 02-MAY-2003  
DEFINITION Bos taurus AMELX gene for amelogenin, partial cds.  
ACCESSION AB091789  
VERSION AB091789.1 GI:29126030  
KEYWORDS  
SOURCE Bos taurus (cow)  
ORGANISM Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.

REFERENCE  
AUTHORS Iwase,M., Satra,Y., Hirai,Y., Hirai,H., Imai,H. and Takahata,N.  
TITLE From the Cover: The amelogenin loci span an ancient pseudautosomal boundary in diverse mammalian species  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 100 (9), 5258-5263 (2003)  
MEDLINE 22608569  
PUBMED 12672962

REFERENCE  
AUTHORS Iwase,M., Satra,Y. and Takahata,N.  
TITLE Direct Submision  
JOURNAL Submitted (19-SEP-2002) Mineyo Iwase, Graduate University for Advanced Studies(Sokendai), Department of Biosystems Science; Shonan kokusaiimura, Hayama, Kanagawa 240-0193, Japan (E-mail: iwase@koryuwo1.soken.ac.jp, Tel:81-468-58-1571, Fax:81-468-58-1544)

## FEATURES

source 1..6451  
Location/Qualifiers  
/organism="Bos taurus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9913"  
/sex="male"  
238..6451  
/gene="AMELX"  
join(238..340,1650..1661)  
/gene="AMELX"  
join(1662..1715,4173..4220,5560..5601,5665..5709,5987..6451)  
/gene="AMELX"  
/codon\_start=1  
/product="amelogenin"



JOURNAL DNA Seq. 6 (5), 307-310 (1996)  
MEDLINE 97142134  
PUBMED 8988368  
REFERENCE 2 (bases 1 to 463)  
AUTHORS Oida, S.  
TITLE Direct Submission  
JOURNAL Submitted (14-JAN-1996) Shinichiro Oida, Tokyo Medical and Dental University, Fac. Dentistry, Dept. Biochemistry, 1-5-45 Yushima, Bunkyo-ku, Tokyo 113, Japan (E-mail: s.oida.odchdent.tmd.ac.jp, Tel:03-5803-5448, Fax:03-5803-0187)  
COMMENT On Dec 15, 1997 this sequence version replaced gi:2662357.  
FEATURES  
Location/Qualifiers  
1..463  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/sub\_species="Balb-C"  
/db\_xref="taxon:10090"  
/chromosome="X"  
/map="XF3-F4"  
/clone="MAMEG-2"  
/cell\_type="Leukocyte"  
/clone\_1lb="EMBL-3 SP6-T7"  
/dev\_stage="adult"  
join(D83063.1:203..258,222..233)  
<1..221  
/number=1  
222..287  
/product="amelogenin"  
/number=2  
288..463  
/number=2  
ORIGIN  
Intron  
Query Match 70.3%; Score 49.2; DB 10; Length 463;  
Best Local Similarity 81.4%; Pred. No. 1.7e-08;  
Matches 57; Conservative 0; Mismatches 13; Indels 0; Gaps 0;  
QY 1 GATTTGTTGCTGCTCTGAGACAGCCTTCAGTATCCCTGAGTAAATTTCCAA 60  
DB 245 GATTTGTTGCTGCTCTGAGACAGCCTTCAGTATCCCTGAGTAAATTTCCAA 304  
QY 61 TTTCATTT 70  
DB 305 TTACTATT 314  
RESULT 5  
BV089295/c 695 bp DNA linear STS 15-OCT-2003  
LOCUS RPAWSEB0001191 Roche Palo Alto Mus musculus STS genomic, sequence  
DEFINITION tagged site.  
ACCESSION BV089295  
VERSION BV089295.1 GI:37666774  
KEYWORDS STS.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 695)  
AUTHORS Uenke, J., Liao, G., Cheng, J., Nguyen, A., Bach, C., Puech, A., McPherson, J.D., Foerzler, D. and Peltz, G.  
TITLE Mus musculus SNPs  
JOURNAL Unpublished (2003)  
COMMENT Contact: Jonathan Uenke  
Roche Palo Alto Genetics and Genomics Department  
Roche Palo Alto  
3431 Hillview Ave, Mailstop S3-1, Palo Alto, CA 94024, USA  
Tel: 6508555807  
Email: Jonathan.Uenke@roche.com  
Primer A: No primer submitted  
Primer B: No primer submitted.  
FEATURES  
Location/Qualifiers

source  
1..695  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/map="X-73613-72919-AL805974.7.1.95826"  
/note="SNPs developed from assay sequences derived from 15 different strains of mice (as of October 1, 2003). Those strains include A/J, A/Heu, 129/Sv, AKR/J, B10.D2-H2/OhnJ, BALB/cByJ, BALB/cJ, C3H/HeJ, C57BL/6J, CAST/Ei, DBA/2J, MRL/MpJ, NZB/BinJ, NZW/Lac, SPRET/Ei."  
<1..>695  
ORIGIN  
STS  
Query Match 70.3%; Score 49.2; DB 11; Length 695;  
Best Local Similarity 81.4%; Pred. No. 1.8e-08;  
Matches 57; Conservative 0; Mismatches 13; Indels 0; Gaps 0;  
QY 1 GATTTGTTGCTGCTCTGAGACAGCCTTCAGTATCCCTGAGTAAATTTCCAA 60  
RESULT 6  
BV097603 695 bp DNA linear STS 14-JAN-2004  
LOCUS RPAWSEB0010113 Roche Palo Alto Mus musculus STS genomic, sequence  
DEFINITION tagged site.  
ACCESSION BV097603  
VERSION BV097603.1 GI:40810679  
KEYWORDS STS.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 695)  
AUTHORS Uenke, J., Liao, G., Cheng, J., Nguyen, A., Bach, C., Puech, A., McPherson, J.D., Foerzler, D. and Peltz, G.  
TITLE Mus musculus SNPs  
JOURNAL Unpublished (2003)  
COMMENT Contact: Jonathan Uenke  
Roche Palo Alto Genetics and Genomics Department  
Roche Palo Alto  
3431 Hillview Ave, Mailstop S3-1, Palo Alto, CA 94024, USA  
Tel: 6508555807  
Email: Jonathan.Uenke@roche.com  
Primer A: No primer submitted with this STS  
Primer B: No primer submitted with this STS.  
FEATURES  
Location/Qualifiers  
1..695  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/map="X-73613-72919-AL805974.7.1.95826"  
/note="SNPs developed from assay sequences derived from 15 different strains of mice (as of October 1, 2003). Those strains include A/J, A/Heu, 129/Sv, AKR/J, B10.D2-H2/OhnJ, BALB/cByJ, BALB/cJ, C3H/HeJ, C57BL/6J, CAST/Ei, DBA/2J, MRL/MpJ, NZB/BinJ, NZW/Lac, SPRET/Ei."  
<1..>695  
ORIGIN  
STS  
Query Match 70.3%; Score 49.2; DB 11; Length 695;  
Best Local Similarity 81.4%; Pred. No. 1.8e-08;  
Matches 57; Conservative 0; Mismatches 13; Indels 0; Gaps 0;  
QY 1 GATTTGTTGCTGCTCTGAGACAGCCTTCAGTATCCCTGAGTAAATTTCCAA 60

```

Db      479 GATTGTTGCTCCGCTGCGAGCAGCTTTTGCATGCTCCCTGAGTAAACATCCCA 420
Qy      61 TTTCATTT 70
      419 TTTACTAATT 410

RESULT 7
LOCUS   AF294397                      9384 bp    DNA       linear   ROD 03-SEP-2000
DEFINITION Mus musculus amelogenin gene, promoter and partial cds.
ACCESSION AF294397
VERSION   AF294397.1 GI:965405
KEYWORDS
SOURCE   Mus musculus (house mouse)
ORGANISM Mus musculus; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS Shead,M.L., Paine,M.L., Luo,W., Zhu,D.H., Yoshida,B., Lei,Y.P.,
          Paine,C.T., Chen,L.S., Burslein,J.M., Jitpukdeeudintira,S.,
          White,S.N. and Bringer,P. Jr.
          Transgene animal model for protein expression and accumulation into
          forming enamel
          Connect. Tissue Res. 38 (1-4), 279-286 (1998)
JOURNAL MEDLINE
PUBMED 20515040
JOURNAL 11063035
TITLE   2 (bases 1 to 9384)
REFERENCE Shead,M.L., Zhu,D.-H., Lei,Y.-P. and Paine,M.L.
          Direct Submission
          Submitted (08-AUG-2000) Dentistry, University of Southern
          California, 2250 Alcazar Street, CSA103, Los Angeles, CA 90033, USA
FEATURES
source
1. .9384
   /organism="Mus musculus"
   /mol_type="genomic DNA"
   /db_xref="taxon:10090"
   /chromosome="X"
1. .8546
   7188..7194
   8547..>8612
   /product="amelogenin"
   /note="enamel matrix protein"
   /codon_start=1
   /product="amelogenin"
   /protein_id="AA010078.1"
   /db_xref="GI:965406"
   /translation="MGTWILFACLLGAFAFAMP"

ORIGIN
Query Match      70.3%; Score 49.2; DB 10; Length 9384;
Best Local Similarity 81.4%; Pred. No. 2.7e-08;
Matches 57; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy      1 GATTTGTTGCTCCCTGCGAGCAGCTTTCAATGCTGAGTAAATTTCCAA 60
      8570 GATTTGTTGCTCCCTGCGAGCAGCTTTGCTATGCTCCGATGAGTAAACATCCA 8629

Db      61 TTTCATTT 70
      419 TTTACTAATT 410

LOCUS   AL805974                      95826 bp    DNA       linear   ROD 29-AUG-2003
DEFINITION Mouse DNA sequence from clone RP23-334F21 on chromosome X, complete
sequence.
ACCESSION AL805974
VERSION   AL805974.8 GI:3436495
KEYWORDS HTG.

```

```

SOURCE
ORGANISM Mus musculus (house mouse)
REFERENCE
AUTHORS Shead,M.L., Paine,M.L., Luo,W., Zhu,D.H., Yoshida,B., Lei,Y.P.,
          Paine,C.T., Chen,L.S., Burslein,J.M., Jitpukdeeudintira,S.,
          White,S.N. and Bringer,P. Jr.
          Transgene animal model for protein expression and accumulation into
          forming enamel
          Connect. Tissue Res. 38 (1-4), 279-286 (1998)
JOURNAL MEDLINE
PUBMED 20515040
JOURNAL 11063035
TITLE   2 (bases 1 to 9384)
REFERENCE Shead,M.L., Zhu,D.-H., Lei,Y.-P. and Paine,M.L.
          Direct Submission
          Submitted (08-AUG-2000) Dentistry, University of Southern
          California, 2250 Alcazar Street, CSA103, Los Angeles, CA 90033, USA
COMMENT
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest, except on the rare
occasion of the clone being a YAC.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-334F21 is
from the RPI-23 Mouse BAC Library
constructed by the group of Pieter de Jong.
For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6.

FEATURES
source
1. .95826
   /organism="Mus musculus"
   /mol_type="genomic DNA"
   /db_xref="taxon:10090"
   /chromosome="X"
   /clone_id="RP23-334F21"
   /clone_lib="RPI-23"

ORIGIN
Query Match      70.3%; Score 49.2; DB 10; Length 95826;
Best Local Similarity 81.4%; Pred. No. 3.7e-08;
Matches 57; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy      1 GATTTGTTGCTCCCTGCGAGCAGCTTTCAATGCTGAGTAAATTTCCAA 60
      22692 GATTTGTTGCTCCCTGCGAGCAGCTTTGCTATGCTCCGATGAGTAAACATCCA 22633

Db      61 TTTCATTT 70
      22632 TTTACTAATT 22623

RESULT 9
LOCUS   AC093946                      212886 bp    DNA       linear   HTG 10-MAY-2003
DEFINITION Ratius norvegicus clone CH230-166F2, *** SEQUENCING IN PROGRESS ***
ACCESSION AC093946

```



VERSION  
 AC093946.9 GI:30520433  
 HTG: HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_ENRICHED.  
 SOURCE  
 Rattus norvegicus (Norway rat)  
 ORGANISM  
 Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 1 (bases 1 to 212886)  
 REFERENCE  
 Muzny, D., Maric, M., Metzker, M., Lee, A., Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Arguliano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blych, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Cessari, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., DeSouza, L., Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Ercotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Frazer, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, M., Guevara, W., Gunaratne, P., Haaland, M., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlik, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogue, M., Hollins, B., Howells, S., Hulik, S., Hume, J., Idelbird, D., Jackson, A., Jackson, J., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowls, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorene, L., Louie, L., Louie, H., Lozano, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapa, P., Martin, K., Martin, R., Martinez, E., Maloney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Mlosoajevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Mundasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwokedi, O., Okunogun, G., Olarnunsgoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfankuch, C., Plopper, F., Polidexter, A., Popovic, D., Primus, E., Pu, L., Pu, M., Putroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rivas, C., Rodkey, T., Rojas, A., Rose, R., Rose, R., Ruiz, S.J., Sanders, W., Saverly, G., Scherer, S., Scott, G., Shattman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajls, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorrelle, R., Soze, J., Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabors, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umanil, K., Valse, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J., Wang, O., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczky, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Zhou, D., von Weizsacker, G., and Gibbs, R.A.  
 TITLE  
 Unpublished  
 JOURNAL  
 2 (bases 1 to 212886)  
 REFERENCE  
 Worley, K.C.  
 TITLE  
 Direct Submission  
 JOURNAL  
 Submitted (12-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 212886)  
 REFERENCE  
 Rat Genome Sequencing Consortium.  
 TITLE  
 Direct Submission  
 JOURNAL  
 Submitted (10-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 COMMENT  
 On May 10, 2003 this sequence version replaced GI:2365261.  
 The sequence in this assembly is a combination of BAC based reads

and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: GPVN  
 Center clone name: CH230-166F2  
 ----- Summary Statistics  
 Assembly program: Atlas 3.0i  
 Consensus quality: 207370 bases at least Q40  
 Consensus quality: 208312 bases at least Q30  
 Consensus quality: 208768 bases at least Q20  
 Estimated insert size: 218822; sum-of-contigs estimation  
 Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

\* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 1 contigs. Gaps between the contigs  
 \* are represented as runs of N. The order of the pieces  
 \* is believed to be correct as given, however the sizes  
 \* of the gaps between them are based on estimates that have  
 \* provided by the submitter.  
 \* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.  
 \* 1 212886: contig of 212886 bp in length.

FEATURES  
 source  
 1. 212886  
 /organism="Rattus norvegicus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10116"  
 /clone="CH230-166F2"  
 1. 1230  
 /note="wgs contig"  
 211449. 212886  
 /note="wgs\_contig"  
 ORIGIN  
 Query Match 68.0%; Score 47.6; DB 2; Length 212886;  
 Best Local Similarity 80.0%; Pred. No. 1.8e-07;  
 Matches 56; Conservative 0; Mismatches 14; Indels 0; Gaps 0;  
 QY 1 GATTTGTTTGGCTGCTCTGAGACAGCTTCAATGAGCTGTGATTAATTTCGA 60  
 DB 112447 GATCTGTGTCCTGCTCTGAGACAGCTTGTGATGAGCTGTGATTAATTTCGA 112388  
 QY 61 TTTCAATT 70  
 DB 112387 TTTACTGATT 112378  
 RESULT 10  
 AC121424 234471 bp DNA 1linear HTG 21-SEP-2002  
 LOCUS  
 DEFINITION  
 Rattus norvegicus clone CH230-122F17, \*\*\* SEQUENCING IN PROGRESS  
 AC121424  
 AC121424  
 AC121424.3 GI:2365350  
 HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.  
 Rattus norvegicus (Norway rat)

## ORGANISM

*Rattus norvegicus*  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.

REFERENCE  
AUTHORS

- 1 (bases 1 to 234471)  
Munzy,D,Marle, Metzker,M,lee, Abramson,S, Adams,C, Alder,J, Allen,C, Allen,H, Alebrooks,S, Amin,A, Anguiano,D, Anyalebechi,V, Aoyagi,A, Ayodeji,M, Baca,E, Baden,H, Baldwin,D, Bandaranaike,D, Barber,M, Barnstead,M, Benahmed,F, Biewald,K, Blair,J, Blankenburg,K, Blyth,P, Brown,M, Bryant,N, Buhay,C, Burch,P, Burrell,K, Calderon,B, Cardenas,V, Carter,K, Cavazos,I, Ceasar,H, Center,A, Chacko,J, Chavez,D, Chen,G, Chen,R, Chen,Y, Chen,Z, Chu,J, Cleveland,C, Cockrell,R, Cox,C, Coyle,M, Cree,A, D'Souza,L, Davila,M,L, Davis,C, Davy-Carroll,L, De Anda,C, Dedrich,D, Delgado,O, Denson,S, Deramo,C, Ding,Y, Dinh,H, Diya,K, Draper,H, Dugan-Rocha,S, Dunn,A, Durbin,K, Duval,B, Eaves,K, Egan,A, Escotto,M, Eugene,C, Evans,C,A, Falls,T, Fan,G, Fernandez,S, Finley,M, Flagg,N, Forbes,L, Foster,M, Foster,P, Fraser,C,M, Gabis,A, Ganta,R, Garcia,A, Garner,T, Garza,M, Gebregeorgis,E, Geer,K, Gill,R, Grady,M, Guerra,M, Guevara,W, Gunaratne,P, Healand,W, Hamli,C, Hamilton,C, Hamilton,K, Harvey,Y, Havlak,P, Hawes,A, Henderson,N, Hernandez,J, Hernandez,R, Hines,S, Hladun,S,L, Hodgson,A, Hogues,M, Hollins,B, Howells,S, Hulik,S, Hume,J, Idlebird,D, Jackson,A, Jackson,L, Jacob,L, Jiang,H, Johnson,B, Johnson,R, Jolivet,A, Karpathy,S, Kelly,S, Kelly,S, Khan,Z, King,L, Kovar,C, Kowis,C, Kraft,C,L, Lebow,H, Lavan,J, Lewis,L, Li,Z, Liu,J, Liu,J, Liu,W, Liu,Y, London,P, Longacre,S, Lopez,O, Lorenshwa,L, Louieged,H, Lozada,R,J, Lu,X, Ma,J, Maheshwari,M, Mahindaratne,M, Mahmoud,M, Malloy,K, Mangum,A, Mangum,B, Mapa,P, Martin,K, Martin,R, Martinez,E, Mawhney,S, McLeod,M,P, McNeill,T,Z, Meenen,E, Mloaavljovic,A, Miner,G, Minja,E, Montemayor,J, Moore,S, Morgan,M, Morris,K, Morris,S, Munidasa,M, Murphy,M, Nait,L, Nankervis,C, Neal,D, Newton,N, Nguyen,N, Norris,B, Nwaokwelen,O, Okunnu,G, Olarnpungoon,A, Pal,S, Parks,K, Pasternak,S, Paul,H, Perez,A, Perez,L, Pfankoch,C, Plopper,F, Poindexter,A, Popovic,D, Primus,E, Pu,L, L, Piaz,M, Quiroz,J, Rachin,E, Reeves,K, Regier,M,A, Reich,R, Reilly,B, Reilly,M, Ren,Y, Reuter,M, Richards,S, Riggs,F, Rives,C, Rodkey,T, Rojas,A, Rose,M, Rose,R, Ruiz,S,J, Sanders,W, Saverly,G, Scherer,S, Scott,G, Shatsman,S, Shan,H, Shetty,U, Shvartsbeyn,A, Sisson,I, Sitter,C, Smays,D, Sneed,A, Sodergren,E, Song,X-Z, Sorelle,R, Sosa,J, Steinle,M, Strong,R, Sutton,A, Svatek,A, Tabors,P, Taylor,C, Taylor,T, Thomas,N, Thomas,S, Tinney,A, Trejos,Z, Uemami,K, Valas,R, Vera,V, Villasana,D, Waldron,L, Walker,B, Wang,J, Wang,Q, Wang,S, Warren,J, Warren,R, Wei,X, White,F, Williams,G, Willson,R, Wleczyk,R, Wooden,H, Worley,K, Wright,D, Wright,R, Wu,J, Yakub,S, Yen,J, Yoon,L, Yoon,V, Yu,F, Zhang,J, Zhou,J, Zhou,X, Zhou,S, Dunn,D, von Niederhausen,A, Weiss,R, Smith,D,R, Holt,R,A, Smith,H,O, Weinstein,G, and Gibbs,R.A.
- Direct Submission  
Unpublished  
2 (bases 1 to 234471)  
Worley,K.C.  
Direct Submission  
Submitted (18-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 234471)  
Rat Genome Sequencing Consortium.  
Rat Genome Sequencing Consortium.  
Submitted (21-SEP-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Sep 21, 2002 this assembly version replaced gi:21908624.  
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). As a result, the sequence may extend beyond the ends of the clone and there may be

contigs that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

## Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

Project Information

Center project name: GXUI

Center clone name: CH230-122F17

Summary Statistics

Assembly program: Phrap, version 0.990329

Consensus quality: 220465 bases at least Q40

Consensus quality: 222988 bases at least Q30

Consensus quality: 224416 bases at least Q20

Estimated insert size: 241282; sum-of-contigs estimation

Quality coverage: 4x in Q20 bases; sum-of-contigs estimation

\* NOTE: Estimated insert size may differ from sequence length

\* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).

\* NOTE: This is a "working draft" sequence. It currently

\* consists of 2 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved.

\* 1 232989: contig of 232989 bp in length

\* 232990 233089: gap of unknown length

\* 233090 234471: contig of 1382 bp in length.

Location/Qualifiers

1. 234471

/organism="Rattus norvegicus"

/mol\_type="genomic DNA"

/db\_xref="taxon:10116"

/clone="CH230-122F17"

104912..105798

/note="clone boundary"

clone\_end:77

site:ECORI

end\_sequence:BH268493"

complement(231872..232742)

/note="clone boundary"

clone\_end:sp6

site:ECORI

end\_sequence:BH268494"

ORIGIN

Query Match 68.0%; Score 47.6; DB 2; Length 234471;

Best Local Similarity 80.0%; Pred. No. 1.8e-071;

Matches 56; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 1 GATTTGTTTCCTGCTCTGAGAGAGCCCTTCAGTATGCTGTGAGTAATTTCCAA 60

DB 161729 GATCTGTTTCCTGCTCTGAGAGAGCCCTTCAGTATGCTGTGAGTAATTTCCAA 161788

QY 61 TTTCCTAATT 70

DB 161789 TTACTGATT 161798

RESULT 11

AB091786

LOCUS AB091786 8004 bp DNA linear PRI 02-MAY-2003

DEFINITION Lemur catta AMELY gene for amelogenin, partial cds.

AB091786

AB091786 1 GI:29126025

KEYWORDS

SOURCE

ORGANISM

Lemur catta (ring-tailed lemur)

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

/organism="Sus scrofa"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9823"  
 /sex="male"  
 1491..7425  
 /gene="AMELX"  
 join(1491..1548,3140..3151)  
 /gene="AMELX"  
 join(3152..3205,5316..5363,6619..6660,6717..6761,  
 7033..7425)  
 /gene="AMELX"  
 /codon\_start=1  
 /product="amelogenin"  
 /protein\_id="BAC6110.1"  
 /db\_xref="GI:29126035"  
 /translation="MGTWILFACILGAAPSPRPHPGHGYNFSEYDLYLEAIRID  
 RTAVSLPLKMYOMNIIPIYTSYGYEPMGMLHHQIIPIVSOOTPGSHALPHNIIEM  
 VPNOQPIPOQPMPIPLPGHSMPTPHOHNLPIPAQPPQPPQVQPPHPIQPPSP  
 MHPQLPLPPPLPMPMSMSLPLDLELMPAT"

ORIGIN

Query Match 65.7%; Score 46; DB 4; Length 7425;  
 Beat Local Similarity 83.9%; Pred. No. 5e-07;  
 Matches 52; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Oy 1 GATTTGTTGCTGCTGCTCCTCGAGAGACGCTTCAGTAGCCTGTGAGTAAATTGCCA 60  
 Db 3163 GATTTGTTGCTGCTGCTCCTCGAGAGACGCTTCAGTAGCCTGTGAGTAAACACCCC 3222  
 Oy 61 TT 62  
 Db 3223 TT 3224

RESULT 13  
 AB091783  
 LOCUS AB091783 6442 bp DNA linear PRI 02-MAY-2003  
 DEFINITION Saimiri sciureus AMELX gene for amelogenin, partial cds.  
 ACCESSION AB091783  
 VERSION AB091783.1 GI:29126019  
 KEYWORDS  
 SOURCE Saimiri sciureus (common squirrel monkey)  
 ORGANISM Saimiri sciureus  
 Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Cebinae;  
 Saimiri.

REFERENCE 1  
 AUTHORS Iwase,M., Saita,Y., Hirai,Y., Hirai,H., Inai,H. and Takahata,N.  
 TITLE From the Cover: The amelogenin loci span an ancient pseudautosomal  
 boundary in diverse mammalian species  
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 100 (9), 5258-5263 (2003)  
 MEDLINE 22608569  
 PUBMED 12672962  
 REFERENCE 2 (bases 1 to 6442)  
 AUTHORS Iwase,M., Saita,Y. and Takahata,N.  
 TITLE Direct Submision  
 JOURNAL Submitted (19-SEP-2002) Mineyo Iwase, Graduate University for  
 Advanced Studies(Sokendai), Department of Biogeorems Science;  
 Shonan kokuuaimura, Hayama, Kanagawa 240-0193, Japan  
 (E-mail:iwase@min@koryuwo1.soken.ac.jp, Tel:81-468-58-1571,  
 Fax:81-468-58-1544)  
 FEATURES  
 source location/Qualifiers  
 1..6442  
 /organism="Saimiri sciureus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9521"  
 /sex="male"  
 936..6442  
 /gene="AMELX"  
 join(936..991,2298..2309)  
 /gene="AMELX"  
 join(2310..2363,4289..4336,5592..5633,5725..5769,  
 6041..6442)

/gene="AMELX"  
/codon\_start=1  
/product="amelogenin"  
/protein\_id="BAC66103.1"  
/db\_xref="GI:29126020"  
/translation="MGTTWILFACLLGAAPFAMPPLPRPHGHPYINSEYSHSOAINID  
RGLVUTPLKQYOSMTIRAPVSGYBPMGWLHQLIIVLSQOHPTHTLQPHHNPV  
VPAQOPVPOQPMVPGQSHMTPTNHQPNLPPLPRAQOPFOPTQVOPQPHQPMQOP  
VHPMOPLPQPLPMPFPMQPLPPLPDLPLEAMPAT"

## ORIGIN

Query Match 64.3%; Score 45; DB 9; Length 6442;  
Best Local Similarity 90.6%; Pred. No. 1.3e-06;  
Matches 48; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

1 GATTTGTTGCTGCTGCTGAGAGAGCTTCAGTATGCTGTGATGAAA 53  
2321 GATTTGTTGCTGCTGCTGAGAGAGCTTTGCCATGCTGTGATGAAA 2373

RESULT 14  
AB091787 5562 bp DNA linear PRI 02-MAY-2003  
LOCUS Ocolemur garnettii AMELX gene for amelogenin, partial cds.  
DEFINITION AB091787  
ACCESSION AB091787  
VERSION AB091787.1 GI:29126027  
KEYWORDS  
SOURCE Ocolemur garnettii (small-eared galago)  
ORGANISM Ocolemur garnettii  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Strepsirhini; Galagonidae; Ocolemur.

REFERENCE  
AUTHORS Iwase, M., Satta, Y., Hirai, Y., Hirai, H., Imai, H. and Takahata, N.  
TITLE From the Cover: The amelogenin loci span an ancient pseudautosomal boundary in diverse mammalian species  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 100 (9), 5258-5263 (2003)  
MEDLINE 22608569  
PUBMED 12672962  
REFERENCE  
AUTHORS Iwase, M., Satta, Y. and Takahata, N.  
TITLE Direct Submission  
JOURNAL Submitted (19-SEP-2002) Mineyo Iwase, Graduate University for Advanced Studies (Sokendai), Department of Biosystems Science; Shonan kokusaiimura, Hayama, Kanagawa 240-0193, Japan  
(E-mail: iwase@minokoryu.w01.soken.ac.jp, Tel: 81-468-58-1571, Fax: 81-468-58-1544)

FEATURES  
source

1..5562  
Location/Qualifiers  
/organism="Ocolemur garnettii"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:30611"  
/sex="male"  
267..5562  
/gene="AMELX"  
join(267..321,1401..1412)  
/gene="AMELX"  
join(1413..1466,3385..3432,4699..4740,4832..4876,  
5125..5562)  
/gene="AMELX"  
/codon\_start=1  
/product="amelogenin"  
/protein\_id="BAC66107.1"  
/db\_xref="GI:29126028"  
/translation="MGTTWILFACLLGAAPFAMPPLPRPHGHPYINSEYSHSOAINID  
RGLVUTPLKQYOSMTIRAPVSGYBPMGWLHQLIIVLSQOHPTHTLQPHHNPV  
VPAQOPVPOQPMVPGQSHMTPTNHQPNLPPLPRAQOPFOPTQVOPQPHQPMQOP  
MOPMOPLPQPLPMPFPMQPLPPLPDLPLEAMPAT"

## ORIGIN

Query Match 62.3%; Score 43.6; DB 9; Length 5562;  
Best Local Similarity 84.5%; Pred. No. 4.5e-06;  
Matches 49; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

1 GATTTGTTGCTGCTGCTGAGAGAGCTTCAGTATGCTGTGATGAAAATTTC 58  
1424 GATTTGTTGCTGCTGCTGAGAGAGCTTTGCCATGCGCCGATGAAAACACC 1481

RESULT 15  
AB091785 5684 bp DNA linear PRI 02-MAY-2003  
LOCUS Lemur catca AMELX gene for amelogenin, partial cds.  
DEFINITION AB091785  
ACCESSION AB091785  
VERSION AB091785.1 GI:29126023  
KEYWORDS  
SOURCE Lemur catca (ring-tailed lemur)  
ORGANISM Lemur catca  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Strepsirhini; Lemniridae; Lemur.

REFERENCE  
AUTHORS Iwase, M., Satta, Y., Hirai, Y., Hirai, H., Imai, H. and Takahata, N.  
TITLE From the Cover: The amelogenin loci span an ancient pseudautosomal boundary in diverse mammalian species  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 100 (9), 5258-5263 (2003)  
MEDLINE 22608569  
PUBMED 12672962  
REFERENCE  
AUTHORS Iwase, M., Satta, Y. and Takahata, N.  
TITLE Direct Submission  
JOURNAL Submitted (19-SEP-2002) Mineyo Iwase, Graduate University for Advanced Studies (Sokendai), Department of Biosystems Science; Shonan kokusaiimura, Hayama, Kanagawa 240-0193, Japan  
(E-mail: iwase@minokoryu.w01.soken.ac.jp, Tel: 81-468-58-1571, Fax: 81-468-58-1544)

FEATURES  
source

1..5684  
Location/Qualifiers  
/organism="Lemur catca"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9447"  
/sex="male"  
100..5684  
/gene="AMELX"  
join(100..156,1503..1514)  
/gene="AMELX"  
join(1515..1568,3465..3512,4804..4845,4937..4981,  
5229..5684)  
/gene="AMELX"  
/codon\_start=1  
/product="amelogenin"  
/protein\_id="BAC66105.1"  
/db\_xref="GI:29126024"  
/translation="MGTTWILFACLLGAAPFAMPPLPRPHGHPYINSEYSHSOAINID  
RGLVUTPLKQYOSMTIRAPVSGYBPMGWLHQLIIVLSQOHPTHTLQPHHNPV  
VPAQOPVPOQPMVPGQSHMTPTNHQPNLPPLPRAQOPFOPTQVOPQPHQPMQOP  
MOPMOPLPQPLPMPFPMQPLPPLPDLPLEAMPAT"

## ORIGIN

Query Match 62.3%; Score 43.6; DB 9; Length 5684;  
Best Local Similarity 84.5%; Pred. No. 4.5e-06;  
Matches 49; Conservative 0; Mismatches 9; Indels 0; Gaps 0;  
1 GATTTGTTGCTGCTGCTGAGAGAGCTTCAGTATGCTGTGATGAAAATTTC 58  
1526 GATTTGTTGCTGCTGCTGAGAGAGCTTTGCCATGCGCCGATGAAAACACC 1583

Search completed: February 8, 2005, 16:36:55  
Job time : 391.297 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 8, 2005, 11:25:10 ; Search time 97.858 Seconds  
(without alignment) 3957.220 Million cell updates/sec

Title: US-10-754-437-7

Perfect score: 70  
Sequence: 1 gatcttgctgcgcgcctcc.....aaattccaatcccaatt 70

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4134886 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 10%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: N\_Geneseq\_23Sep04:\*  
2: geneseqn1980s:\*  
3: geneseqn1990s:\*  
4: geneseqn2000s:\*  
5: geneseqn2001as:\*  
6: geneseqn2001bs:\*  
7: geneseqn2002as:\*  
8: geneseqn2002bs:\*  
9: geneseqn2003as:\*  
10: geneseqn2003bs:\*  
11: geneseqn2003ds:\*  
12: geneseqn2004s:\*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	43.6	62.3	259202	12	ADQ18492 Human sof
2	40.2	57.4	722	6	AA141111 GAML, rela
3	40.2	57.4	752	6	AA141110 GAML, rela
4	38	54.3	270	12	ADM80835 Human CAD
5	38	54.3	549	12	ADM80826 Human CAD
6	38	54.3	623	12	ADM80827 Human CAD
7	35.6	50.9	825	4	AAH03775 Human CDN
8	35.6	50.9	1626	4	AAH17575 Human CDN
9	35.6	50.9	2176	4	AAH17101 Human CDN
10	35.6	50.9	3037	12	ADQ24117 Human sof
11	34.6	49.4	556	10	ADM59026 Toxicity
12	34.6	49.4	556	10	ADM53782 Primary r
13	34	48.6	65	6	ABN28389 Rat splic
14	34	48.6	476	6	AAZ50832 Rat amelo
15	33.8	48.3	318	12	ADM80836 Human CAD
16	33.8	48.3	750	2	AAZ07020 Amelogeni
17	33.8	48.3	793	2	AAZ07018 Amelogeni
18	32.2	46.0	802	2	AAZ07019 Amelogeni
19	32.2	46.0	852	12	ADQ22977 Human sof
20	28.6	40.9	277	3	AAC55907 Eucalyptu
21	26.4	37.7	389	5	AAFe4393 Novel hum

C 22	26.4	37.7	423	4	ABA58214 Human foe
C 23	26.4	37.7	423	4	AAI37826 Probe #65
C 24	26.4	37.7	423	4	AAK31962 Human bon
C 25	26.4	37.7	423	4	AAK06298 Human bra
C 26	26.4	37.7	423	4	AB831653 Human liv
C 27	26.4	37.7	423	6	AB806725 Human gen
C 28	26.4	37.7	567	4	AA143405 Probe #12
C 29	26.4	37.7	941	2	AAV84413 Human sec
C 30	26.4	37.7	941	2	AAV84618 Human sec
C 31	26.4	37.7	941	4	ABA83196 Human sec
C 32	26.4	37.7	941	4	ABA83401 Human sec
C 33	26.4	37.7	941	9	ACH04902 Novel hum
C 34	26.4	37.7	941	9	ACH04697 Novel hum
C 35	26.4	37.7	941	9	ACD44712 Human CDN
C 36	26.4	37.7	941	9	ACD44507 Human CDN
C 37	25.8	36.9	292	3	ACC28078 Human sec
C 38	25.8	36.9	2000	10	ACC61018 Gene sequ
C 39	25.8	36.9	2000	10	ADK62663 Disease t
C 40	25.4	36.3	199878	10	ADL13719 Osteonarth
C 41	25.2	36.0	1639	12	ADN05617 Antipsocti
C 42	25.2	36.0	48436	6	ABN89533 Human cor
C 43	25.2	36.0	96588	9	ADA03026 Human MEN
C 44	25.2	36.0	96588	10	ADB72764 Human MEN
C 45	25.2	36.0	96588	10	ADC85506 Human Mbn

#### ALIGNMENTS

RESULT 1	ADQ18492	ADQ18492 standard; DNA; 259202 BP.
AC	ADQ18492;	
DT	26-AUG-2004 (first entry)	
DB	Human soft tissue sarcoma-upregulated DNA - SEQ ID 1311.	
KW	soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human; ds.	
OS	Homo sapiens.	
XX	10-JUN-2004.	
XX	26-NOV-2003; 2003WO-US038193.	
XX	26-NOV-2002; 2002US-0429739P.	
XX	(PROT-) PROTEIN DESIGN LABS INC.	
XX	Aziz N, Ginsburg WM, Zlotnik A;	
XX	WPI; 2004-441208/41.	
XX	Early detection of soft tissue sarcoma comprises determining expression of a gene in a first soft tissue sample and a normal soft tissue sample and comparing the gene expression, also useful in treating soft tissue sarcoma.	
XX	Example 2; SEQ ID NO 1311; 210bp; English.	
XX	The invention relates to a novel method for detecting soft tissue sarcoma which comprises obtaining a first soft tissue sample from an individual, and a normal soft tissue sample from the same or different individual, determining the expression of a gene in both samples and comparing the expression of the gene in both soft tissue samples, where a higher level of protein expression in the first soft tissue sample indicates the presence of soft tissue sarcoma. The method of the invention has cytostatic applications and may be useful for detecting soft tissue	

CC sarcoma, possibly via gene therapy or vaccine production. The nucleic  
CC acid sequences may be useful in diagnostic and screening applications.  
CC The current sequence is that of a human soft tissue sarcoma-upregulated  
CC DNA of the invention. The current sequence is not shown within the  
CC specification per se but was submitted in CD format by the inventor.

XX Sequence 259202 BP; 81699 A; 51421 C; 49221 G; 76861 T; 0 U; 0 Other;

CC Query Match 62.3%; Score 43.6; DB 12; Length 259202;

XX Best Local Similarity 84.5%; Pred. No. 2.4e-05;

XX Matches 49; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

CC 1 GATTTTGTTCCTGCTGCTGAGAGACCTTCACTATGCGGTGAGTAAATTTCC 58

XX DB 69884 GATTTTATTTGCTGCTGCTGAGACCTTTTGCATGCTGTGATTAACACCC 69941

RESULT 2

XX ID AAL41111 standard; DNA; 722 BP.

XX AAL41111;

XX 16-OCT-2002 (first entry)

XX GAML related Y-chromosome DNA sequence.

XX Goat embryo sexual identification technique; goat amelogenin gene; GAML;

XX sex-specific; gene; ds; Y-chromosome.

XX Capra hircus.

XX Key Location/Qualifiers

XX CDS 35..658

XX /tag= a

XX /product= "Y-chromosome protein"

XX TW454013-A.

XX 11-SEP-2001.

XX 10-NOV-1999; 99TW-00119616.

XX 10-NOV-1999; 99TW-00119616.

XX (CHEN/) CHEN C.

XX (JANG/) JANG J.

XX (WENG/) WENG T.

XX (JENG/) JENG D.

XX Chen C, Jang J, Weng T, Jeng D;

XX WPI; 2002-442016/47.

XX P-PSDB; AAO22534.

XX Sex-specific sequence of goat amelogenin gene, useful for embryo sexual

XX identification, comprises high sensitivity even using single white blood

XX cell or cleavage c.

XX Disclosure; Page 28; 35pp; Chinese.

XX The invention relates to a goat embryo sexual identification technique

XX with high efficiency, sensitivity and repeatability. This technique

XX involves separately cloning and sequencing the coding regions and the

XX introns of the goat amelogenin gene (GAML) on the goat chromosomes. The

XX results indicate that there are sex-specific sequences in the fifth

XX intron of the gene. The major characteristics according to the present

XX invention include high sensitivity, applicable in sex identification even

XX only using a single white blood cell or a single cleavage cell of

XX blastula; high diagnostic efficiency, capable of identifying hundreds of

XX goat embryo in 3 hours; simple operation procedures without complicated

XX steps of DNA extraction and need no additional control group intron; and

XX can be applied on different species of goats. This polynucleotide

CC sequence represents a GAML related Y-chromosome DNA sequence of the

XX invention

XX Sequence 722 BP; 173 A; 264 C; 140 G; 145 T; 0 U; 0 Other;

XX Query Match 57.4%; Score 40.2; DB 6; Length 722;

XX Best Local Similarity 93.3%; Pred. No. 7.3e-05;

XX Matches 42; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

CC 1 GATTTTGTTCCTGCTGCTGAGAGACCTTCACTATGCTGT 45

XX DB 46 GATTTTGTTCCTGCTGCTGAGAGACCTTCACTATGCTGT 90

RESULT 3

XX ID AAL41110 standard; DNA; 752 BP.

XX AAL41110;

XX 16-OCT-2002 (first entry)

XX GAML related X-chromosome DNA sequence.

XX Goat embryo sexual identification technique; goat amelogenin gene; GAML;

XX sex-specific; gene; ds; X-chromosome.

XX Capra hircus.

XX Key Location/Qualifiers

XX CDS 35..658

XX /tag= a

XX /product= "X-chromosome protein"

XX TW454013-A.

XX 11-SEP-2001.

XX 10-NOV-1999; 99TW-00119616.

XX 10-NOV-1999; 99TW-00119616.

XX (CHEN/) CHEN C.

XX (JANG/) JANG J.

XX (WENG/) WENG T.

XX (JENG/) JENG D.

XX Chen C, Jang J, Weng T, Jeng D;

XX WPI; 2002-442016/47.

XX P-PSDB; AAO22534.

XX Sex-specific sequence of goat amelogenin gene, useful for embryo sexual

XX identification, comprises high sensitivity even using single white blood

XX cell or cleavage c.

XX Disclosure; Page 28; 35pp; Chinese.

XX The invention relates to a goat embryo sexual identification technique

XX with high efficiency, sensitivity and repeatability. This technique

XX involves separately cloning and sequencing the coding regions and the

XX introns of the goat amelogenin gene (GAML) on the goat chromosomes. The

XX results indicate that there are sex-specific sequences in the fifth

XX intron of the gene. The major characteristics according to the present

XX invention include high sensitivity, applicable in sex identification even

XX only using a single white blood cell or a single cleavage cell of

XX blastula; high diagnostic efficiency, capable of identifying hundreds of

XX goat embryo in 3 hours; simple operation procedures without complicated

XX steps of DNA extraction and need no additional control group intron; and

XX can be applied on different species of goats. This polynucleotide

XX sequence represents a GAML related X-chromosome DNA sequence of the

XX invention

**SQ** Sequence 270 BP; 76 A; 65 C;

neurological, developmental,

disorders including cancer, e.g. breast, prostate, ovarian, lung or colon



CC cancer, obesity and Tangier disease.  
XX  
SQ Sequence 549 BP; 134 A; 203 C; 112 G; 100 T; 0 U; 0 Other;  
Query Match 54.3%; Score 38; DB 12; Length 549;  
Best Local Similarity 75.8%; Pred. No. 0.00045;  
Matches 47; Conservative 0; Mismatches 15; Indels 0; Gaps 0;  
XX  
QY 1 GATTTGTTGGCTGCTCTCTGGAGACGCTTCAGTATGCTGTGAGTAAATTTCGA 60  
DB 47 GATTTATTTGCTGCTCTCTGGAGACGCTTCAGTATGCTGTGAGTAAACCCTTTGA 106  
QY 61 TT 62  
DB 107 GT 108  
RESULT 6  
ADM80827  
ID ADM80827 standard; cDNA; 623 BP.  
XX  
AC ADM80827;  
XX  
DT 03-JUN-2004 (first entry)  
XX  
DE Human CADEC14 encoding cDNA SEQ ID NO:56.  
XX  
XX human; cell adhesion and extracellular matrix protein; CADEC14;  
XX neuroprotective; cytoskeletal; anorectic; immune disorder;  
XX neurological disorder; developmental disorder;  
XX connective tissue disorder; cell proliferative disorder; cancer; obesity;  
XX Tangier disease; gene; ss.  
XX  
XX Homo sapiens.  
XX  
OS  
XX  
FH Key Location/Qualifiers  
FT CDS 36..563  
FT FT /\*tag= a  
FT FT /product= "CADEC14"  
XX  
XX MO2004015396-A2.  
XX  
XX 19-FEB-2004.  
XX  
XX 12-AUG-2003; 2003WO-US025418.  
XX  
XX 13-AUG-2002; 2002US-0403781P.  
XX 30-AUG-2002; 2002US-0407034P.  
XX 13-SEP-2002; 2002US-0410566P.  
XX 24-SEP-2002; 2002US-0413482P.  
XX 25-SEP-2002; 2002US-0413890P.  
XX 08-NOV-2002; 2002US-0424904P.  
XX 13-NOV-2002; 2002US-0426222P.  
XX  
XX (INCY-) INCYTE CORP.  
XX  
XX ELLIOTT VS, Khare R, Emerling BM, Kable AE, Tran UK, Jin P;  
XX Becha SD, Margulis JP, Swarnakar A, Chawla NK, Ramkumar J, Blake JJ;  
XX Hafalia AJA, Lee SY, Jiang X, Jackson AA, Richardson TW, Wang UT;  
XX Wang UT, Chien D, Yang YG;  
XX  
XX WPI; 2004-191795/18.  
XX P-PSDB; ADM80785.  
XX  
XX New cell adhesion and extracellular matrix proteins, useful in  
XX diagnosing, treating and preventing immune, neurological, developmental,  
XX connective tissue and cell proliferative disorders including cancer.  
XX  
XX Claim 5; SEQ ID NO 56; 272pp; English.  
XX  
XX The present sequence encodes a human cell adhesion and extracellular  
XX matrix protein designated CADEC14. CADEC14 sequences has neuroprotective,  
XX cytoskeletal and anorectic activities. The CADEC14 polypeptides and

CC polynucleotides are useful in diagnosing, treating and preventing immune,  
XX neurological, developmental, connective tissue and cell proliferative  
XX disorders including cancer, e.g. breast, prostate, ovarian, lung or colon  
XX cancer, obesity and Tangier disease.  
XX  
SQ Sequence 623 BP; 150 A; 228 C; 128 G; 117 T; 0 U; 0 Other;  
Query Match 54.3%; Score 38; DB 12; Length 623;  
Best Local Similarity 75.8%; Pred. No. 0.00047;  
Matches 47; Conservative 0; Mismatches 15; Indels 0; Gaps 0;  
XX  
QY 1 GATTTGTTGGCTGCTCTCTGGAGACGCTTCAGTATGCTGTGAGTAAATTTCGA 60  
DB 47 GATTTATTTGCTGCTCTCTGGAGACGCTTCAGTATGCTGTGAGTAAACCCTTTGA 106  
QY 61 TT 62  
DB 107 GT 108  
RESULT 7  
AAH03775/C  
ID AAH03775 standard; cDNA; 825 BP.  
XX  
XX AAH03775;  
XX  
DT 26-JUN-2001 (first entry)  
XX  
XX Human CDNA clone (5'-primer) SEQ ID NO:610.  
XX  
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
XX  
XX Homo sapiens.  
XX  
OS  
XX  
FH Key Location/Qualifiers  
FT CDS 28..563  
FT FT /\*tag= a  
FT FT /product= "CADEC14"  
XX  
XX 28-UTL-2000; 2000EP-00116126.  
XX  
XX 29-UTL-1999; 99UP-00248036.  
XX 27-AUG-1999; 99UP-00300253.  
XX 11-JAN-2000; 2000JP-00118776.  
XX 02-MAY-2000; 2000JP-00183767.  
XX 09-JUN-2000; 2000JP-00241899.  
XX  
XX (HELI-) HELIX RES INST.  
XX  
XX Oka T, Isegai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX  
XX WPI; 2001-318749/34.  
XX  
XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-  
XX length CDNA defined in the specification, and for the detection and/or  
XX diagnosis of the abnormality of the proteins encoded by the full-length  
XX CDNA.  
XX  
XX Claim 1; SEQ ID NO 610; 2537pp + Sequence Listing; English.  
XX  
XX The present invention describes primer sets for synthesizing 5602 full-  
XX length CDNA defined in the specification. Where a primer set comprises:  
XX (a) an oligo-dT primer and an oligonucleotide complementary to the  
XX complementary strand of a polynucleotide which comprises one of the 5602  
XX nucleotide sequences defined in the specification; where the  
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
XX of an oligonucleotide comprising a sequence complementary to the  
XX complementary strand of a polynucleotide which comprises a 5'-end  
XX sequence and an oligonucleotide comprising a sequence complementary to a  
XX polynucleotide which comprises a 3'-end sequence; where the  
XX oligonucleotide comprises at least 15 nucleotides and the combination of  
XX the 5'-end sequence/3'-end sequence is selected from those defined in the  
XX specification. The primer sets can be used in antisense therapy and in



CC gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialized methods. AAH03166 to AAH13628 and CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention.

CC Sequence 825 BP; 281 A; 123 C; 163 G; 255 T; 0 U; 3 Other;

Query Match 50.9%; Score 35.6; DB 4; Length 825;

Best Local Similarity 71.2%; Pred. No. 0.0041; Mismatches 19; Indels 0; Gaps 0;

Matches 47; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 4 TTGTTGCTGCTGCTGCTGAGAGACCTTCAGTATGCTGTGAGTAAATTTCCAAATT 63  
DB 458 TTTCCTTGACTGCTCATGAGAGCTAAATTCAGTTTCTGTATTATTAATTACACTTT 399  
QY 64 CCAATT 69  
DB 398 CTCCTT 393

## RESULT 8

AAH17575/c  
ID AAH17575 standard; cDNA; 1626 BP.

AAH17575;

26-JUN-2001 (first entry)

Human cDNA sequence SEQ ID NO:17065.

Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

Homo sapiens.

EP1074617-A2.

07-FEB-2001.

28-JUL-2000; 2000EP-00116126.

29-JUL-1999; 99JP-00248036.

27-AUG-1999; 99JP-00300253.

11-JAN-2000; 2000JP-00118776.

02-MAY-2000; 2000JP-00183767.

09-JUN-2000; 2000JP-00241899.

(HELI-) HELIX RES INST.

Ota T, Isegai T, Nishikawa T, Hayaashi K, Saito K, Yamamoto J;

Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

WPI; 2001-318749/34.

Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.

Claim 8; SEQ ID NO 17065; 2537pp + Sequence Listing; English.  
The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end

CC sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialized methods. AAH03166 to AAH13628 and CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention.

CC Sequence 1626 BP; 509 A; 297 C; 299 G; 521 T; 0 U; 0 Other;

Query Match 50.9%; Score 35.6; DB 4; Length 1626;

Best Local Similarity 71.2%; Pred. No. 0.005; Mismatches 19; Indels 0; Gaps 0;

Matches 47; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 4 TTGTTGCTGCTGCTGCTGAGAGACCTTCAGTATGCTGTGAGTAAATTTCCAAATT 63  
DB 913 TTTCCTTGACTGCTCATGAGAGCTAAATTCAGTTTCTGTATTATTAATTACACTTT 854  
QY 64 CCAATT 69  
DB 853 CTCCTT 848

## RESULT 9

AAH17101/c  
ID AAH17101 standard; cDNA; 2176 BP.

AAH17101;

26-JUN-2001 (first entry)

Human cDNA sequence SEQ ID NO:16427.

Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

Homo sapiens.

EP1074617-A2.

07-FEB-2001.

28-JUL-2000; 2000EP-00116126.

29-JUL-1999; 99JP-00248036.

27-AUG-1999; 99JP-00300253.

11-JAN-2000; 2000JP-00118776.

02-MAY-2000; 2000JP-00183767.

09-JUN-2000; 2000JP-00241899.

(HELI-) HELIX RES INST.

Ota T, Isegai T, Nishikawa T, Hayaashi K, Saito K, Yamamoto J;

Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

WPI; 2001-318749/34.

Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.

Claim 8; SEQ ID NO 16427; 2537pp + Sequence Listing; English.  
The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the

CC complementary strand of a polynucleotide which comprises one of the 5602  
CC nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in the  
CC specification. The primer sets can be used in antisense therapy and in  
CC gene therapy. The primers are useful for synthesizing polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893  
CC represent human amino acid sequences; and AAH13629 to AAH13632 represent  
CC oligonucleotides, all of which are used in the exemplification of the  
CC present invention

SEQ Sequence 2176 BP; 683 A; 445 C; 394 G; 654 T; 0 U; 0 Other;

Query Match 50.9%; Score 35.6; DB 4; Length 2176;  
Best Local Similarity 71.2%; Pred. No. 0.0055;  
Matches 47; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 4 TTTGTTTGCCCTGCGGAGCAGCCCTTCAGTATGCTGAGTAAATTTCCAAATT 63  
DB 458 TTTCCTTACGCTCATGAGAGCTAAATTCAGTTTCTGTATTATACATTACACTTT 399  
QY 64 CCAATT 69  
DB 398 CTCCTT 393

RESULT 10  
ADO24117/C  
ID ADO24117 standard; DNA; 3037 BP.  
XX ADQ24117;  
AC ADQ24117;  
XX 26-AUG-2004 (first entry)  
DT 26-AUG-2004 (first entry)  
XX Human soft tissue sarcoma-upregulated DNA - SEQ ID 6937.  
DE soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;  
KM de.  
XX Homo sapiens.  
OS  
XX WO2004048938-A2.  
PN 10-JUN-2004.  
PD 26-NOV-2003; 2003WO-US038193.  
PF 26-NOV-2003; 2003WO-US038193.  
XX 26-NOV-2002; 2002US-0429739P.  
PR (PROT-) PROTEIN DESIGN LABS INC.  
XX Aziz N, Ginsburg WM, Zlotnick A;  
PI WPI; 2004-441208/41.  
DR  
XX Early detection of soft tissue sarcoma comprises determining expression  
PT of a gene in a first soft tissue sample and a normal soft tissue sample  
PT and comparing the gene expression, also useful in treating soft tissue  
PT sarcoma.  
XX Example 2; SEQ ID NO 6937; 210pp; English.  
PS The invention relates to a novel method for detecting soft tissue sarcoma  
CC which comprises obtaining a first soft tissue sample from an individual

CC and a normal soft tissue sample from the same or different individual,  
CC determining the expression of a gene in both samples and comparing the  
CC expression of the gene in both soft tissue samples, where a higher level  
CC of protein expression in the first soft tissue sample indicates the  
CC presence of soft tissue sarcoma. The method of the invention has  
CC cytostatic applications and may be useful for detecting soft tissue  
CC sarcoma, possibly via gene therapy or vaccine production. The nucleic  
CC acid sequences may be useful in diagnostic and screening applications.  
CC The current sequence is that of a human soft tissue sarcoma-upregulated  
CC DNA of the invention. The current sequence is not shown within the  
CC specification per se but was submitted in CD format by the inventor.

SEQ Sequence 3037 BP; 957 A; 596 C; 545 G; 919 T; 0 U; 20 Other;

Query Match 50.9%; Score 35.6; DB 12; Length 3037;  
Best Local Similarity 71.2%; Pred. No. 0.0061;  
Matches 47; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 4 TTTGTTTGCCCTGCGGAGCAGCCCTTCAGTATGCTGAGTAAATTTCCAAATT 63  
DB 913 TTTCCTTACGCTCATGAGAGCTAAATTCAGTTTCTGTATTATACATTACACTTT 854  
QY 64 CCAATT 69  
DB 853 CTCCTT 848

RESULT 11  
ADB59026  
ID ADB59026 standard; DNA; 556 BP.  
XX ADB59026;  
AC ADB59026;  
XX 04-DEC-2003 (first entry)  
DT 04-DEC-2003 (first entry)  
XX Toxicity-related gene, SEQ ID 4052.  
DE Toxicity-related gene, SEQ ID 4052.  
XX Toxic; toxin; gene expression profile; hepatotoxicity; liver;  
KM drug screening; toxicity assay; de.  
XX Unidentified.  
OS  
XX WO2003064624-A2.  
PN 07-AUG-2003.  
PD 31-JAN-2003; 2003WO-US003194.  
PF 31-JAN-2003; 2003WO-US003194.  
XX 31-JAN-2002; 2002US-00060087.  
PR 15-MAR-2002; 2002US-0364045P.  
XX 15-MAR-2002; 2002US-0364055P.  
PR 30-DEC-2002; 2002US-0436643P.  
XX (GENE-) GENE LOGIC INC.  
PA Mendrick D, Porter M, Johnson K, Higgs B, Castle A, Elashoff M;  
PI WPI; 2003-689530/65.  
DR  
XX Predicting a toxic effect of a compound, useful in identifying toxicity  
PT markers in liver tissues or cells for drug screening and toxicity assays,  
PT comprises preparing gene expression profile of tissue or cells exposed to  
PT the compound.

PS Claim 1; SEQ ID NO 4052; 1156pp; English.  
XX The present invention relates to a method for predicting a toxic effect  
CC of a compound. The method comprises preparing a gene expression profile  
CC of a tissue or cell sample exposed to the compound, and comparing the  
CC gene expression profile to a database comprising SEQ ID 1-4925, where  
CC differential expression of the gene indicates at least one toxic effect.  
CC The method is useful for predicting at least one toxic effect of a  
CC compound, predicting hepatotoxicity or the progression of a toxic effect

of a compound, identifying an agent that modulates the onset or progression of a toxic response, predicting the cellular pathways that a compound modulates in a cell, and identifying an agent that modulates at least one activity of a protein. The method and compositions of the present invention using a database of genes having liver toxin-induced differential expression, are useful in identifying toxicity markers in liver tissues or cells for drug screening and toxicity assays. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 556 BP; 121 A; 183 C; 114 G; 119 T; 0 U; 19 Other;

XX Query Match 49.4%; Score 34.6; DB 10; Length 556;

XX Best Local Similarity 88.1%; Pred. No. 0.0086;

XX Matches 37; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 GATTTGTTGGCTGCTCTCTGGAGACAGCTTCACTATGCC 42

DB 33 GATCTTGTGGCTGCTCTCTGGAGACAGCTTTGNTATGCC 74

RESULT 12

ADB53782

ID ADB53782 standard; DNA; 556 BP.

XX ADB53782;

XX 04-DEC-2003 (first entry)

XX Primary rat hepatocyte toxicity modelling related gene SEQ ID NO:4324.

XX toxic effect; gene expression profile; hepatotoxicity; diagnostic marker;

XX toxicity marker; toxicity progression; drug screening;

XX primary rat hepatocyte toxicity modelling; gene; ds.

XX Rattus norvegicus.

XX WO2003065993-A2.

XX 14-AUG-2003.

XX 04-FEB-2003; 2003WO-US003482.

XX 04-FEB-2002; 2002US-0353171P.

XX 13-MAR-2002; 2002US-0363534P.

XX 08-APR-2002; 2002US-0370248P.

XX 10-APR-2002; 2002US-0371134P.

XX 10-APR-2002; 2002US-0371135P.

XX 10-APR-2002; 2002US-0371150P.

XX 11-APR-2002; 2002US-0371413P.

XX 19-APR-2002; 2002US-0373601P.

XX 22-APR-2002; 2002US-0373602P.

XX 08-MAY-2002; 2002US-0378370P.

XX 09-MAY-2002; 2002US-0378652P.

XX 09-MAY-2002; 2002US-0378653P.

XX 09-MAY-2002; 2002US-0378655P.

XX 09-JUL-2002; 2002US-0394230P.

XX 09-JUL-2002; 2002US-0394253P.

XX 04-SEP-2002; 2002US-0407688P.

XX 28-JAN-2003; 2003US-0442900P.

XX (GENE-) GENE LOGIC INC.

XX Mendrick D, Porter M, Johnson K, Higgs B, Castle A, Orr M;

XX Elashoff M;

XX WPI, 2003-731472/69.

XX Determining if a compound induces a toxic effect on a tissue or cell, for

XX identifying hepatotoxic compounds, comprises comparing a gene expression

XX profile of a tissue or cell sample to a database of Tox mean and non-Tox

PT mean values.

XX Claim 44; SEQ ID NO 4324; 874pp; English.

XX The present invention describes a method for determining whether a

XX compound induces a toxic effect on a tissue or cell. The method comprises

XX preparing a gene expression profile of a tissue or cell sample exposed to

XX the compound, and comparing the gene expression profile to a database

XX comprising data or information on the Tox mean and non-Tox mean value.

XX The method is useful for predicting or identifying at least one toxic

XX effect, particularly hepatotoxicity, of a test or unknown compound. The

XX genes listed in the specification are useful as diagnostic or toxicity

XX markers for the prediction or identification of the physiological state

XX of tissue or cell sample that has been exposed to a compound, or to

XX identify or predict the toxic effects of a compound or an agent. These

XX may also be used as markers for monitoring toxicity progression or for

XX drug screening. The present sequence represents a primary rat hepatocyte

XX toxicity modelling related gene sequence from the present invention.

XX SQ Sequence 556 BP; 121 A; 183 C; 114 G; 119 T; 0 U; 19 Other;

XX Query Match 49.4%; Score 34.6; DB 10; Length 556;

XX Best Local Similarity 88.1%; Pred. No. 0.0086;

XX Matches 37; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 GATTTGTTGGCTGCTCTCTGGAGACAGCTTCACTATGCC 42

DB 33 GATCTTGTGGCTGCTCTCTGGAGACAGCTTTGNTATGCC 74

RESULT 13

ABN28389

ID ABN28389 standard; DNA; 65 BP.

XX AC ABN28389;

XX 15-JUL-2002 (first entry)

XX Rat spliced transcript detection oligonucleotide SEQ ID NO:1137.

XX Human; mouse; rat; splice transcript; detection; RNA transcript;

XX splice variant; transcriptome; oligonucleotide library; ss.

XX Rattus norvegicus.

XX WO200210449-A2.

XX 07-FEB-2002.

XX 20-JUL-2001; 2001WO-IB001903.

XX 28-JUL-2000; 2000US-0221607P.

XX 02-MAY-2001; 2001US-0287724P.

XX (COMP-) COMPUGEN INC.

XX Shoshan A, Wasserman A, Mintz E, Mintz L, Faigler S;

XX WPI, 2002-257383/30.

XX New oligonucleotide libraries comprising oligonucleotides which

XX selectively hybridize to mRNAs transcribed from a transcription unit of a

XX genome, useful for detecting tissue-, pathology-, and developmental-

XX specific genes.

XX Example 1; SEQ ID NO 1137; 47pp; English.

XX The present invention describes oligonucleotide libraries for detecting

XX messenger RNAs that populate a (sub-)transcriptome, where the (sub-

XX )transcriptome comprises messenger RNAs transcribed from multiple

XX transcription units that populate a genome. The library comprises several

XX oligonucleotides, each capable of hybridizing selectively to a set of

XX messenger RNAs transcribed from a given transcription unit of the genome,



```

FT CDS 36..260
PT /*tag= a
PT /product= "CADECM-23"
XX
XX WO2004015396-A2.
XX
XX 19-FEB-2004.
XX
XX 12-AUG-2003; 2003WO-US025418.
XX
XX 13-AUG-2002; 2002US-0403781P.
XX 30-AUG-2002; 2002US-0407034P.
XX 13-SEP-2002; 2002US-0410566P.
XX 24-SEP-2002; 2002US-0413482P.
XX 25-SEP-2002; 2002US-0413890P.
XX 08-NOV-2002; 2002US-0424504P.
XX 13-NOV-2002; 2002US-0426222P.
XX
XX (INCY-) INCYTE CORP.
XX
XX Eliott VS, Khare R, Emerling BM, Kable AE, Tran UK, Jin P,
XX Becha SD, Margulis JP, Swarnakar A, Chavla NK, Ramkumar J,
XX Hafalia AJA, Lee SY, Jiang X, Jackson AA, Richardson TW, Blake JJ,
XX Wang JT, Chien D, Yang YG;
XX
XX WPI; 2004-191795/18.
XX P-PSDB; ADM80794.
XX
XX New cell adhesion and extracellular matrix proteins, useful in
XX diagnosing, treating and preventing immune, neurological, developmental,
XX connective tissue and cell proliferative disorders including cancer.
XX
XX Claim 5; SEQ ID NO 65; 272pp; English.
XX
XX The present sequence encodes a human cell adhesion and extracellular
XX matrix protein designated CADECM. CADECM sequences has neuroprotective,
XX cytostatic and anorectic activities. The CADECM polypeptides and
XX polynucleotides are useful in diagnosing, treating and preventing immune,
XX neurological, developmental, connective tissue and cell proliferative
XX disorders including cancer, e.g. breast, prostate, ovarian, lung or colon
XX cancer, obesity and Tangier disease.
XX
XX Sequence 318 BP; 87 A; 81 C; 75 G; 75 T; 0 U; 0 Other;
SQ

```

Query Match 48.3%; Score 33.8; DB 12; Length 318;  
 Best Local Similarity 84.4%; Pred. No. 0.014;  
 Matches 38; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```

Qy 1 GATTTGTTTGCTGCTGCTGAGAGAGCCTTCAGTAGCCCTGT 45
   |||||
Db 47 GATTTATTGCTGCTGCTGAGAGAGCCTTCAGTAGCCCTCT 91

```

Search completed: February 8, 2005, 13:57:04  
 Job time : 96.858 secs

**This Page Blank (uspto)**



ZIP: 94304-1050  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch diskette  
COMPUTER: IBM compatible  
OPERATING SYSTEM: Microsoft Windows 95  
SOFTWARE: Wordperfect for windows 6.0,  
SOFTWARE: ASCII (DOS) TEXT format  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/018,595B  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
ATTORNEY/AGENT INFORMATION:  
NAME: David J. Weitz  
REGISTRATION NUMBER: 38,362  
REFERENCE/DOCKET NUMBER: PELM-744  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 493-9300  
TELEFAX: (650) 493-6811  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 500 nucleotides  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-09-018-595B-3

Query Match 48.3%; Score 33.8; DB 2; Length 500;  
Best Local Similarity 84.4%; Pred. No. 0.0013;  
Matches 38; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 GATTTGTTGGCTGCTCTGGAGAGAGCTTTCAGTATGCTGT 45  
Db 80 GATTTATTTGCTGCTCTGAGAGAGAGCTTTGCGATGCTCT 124

RESULT 3  
US-09-324-709A-3  
Sequence 3, Application US/09324709A  
Patent No. 6154707  
GENERAL INFORMATION:  
APPLICANT: Perkin-Elmer Corporation,  
APPLICANT: Applied Biosystems Division  
TITLE OF INVENTION: FLUORESCENCE GENOTYPING AT MULTIPLE ALLELIC  
TITLE OF INVENTION: SITES  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David J. Weitz,  
ADDRESSEE: Wilson Sonsini Goodrich & Rosati  
STREET: 650 Page Mill Road  
CITY: Palo Alto  
STATE: California  
COUNTRY: USA  
ZIP: 94304-1050  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch diskette  
COMPUTER: IBM compatible  
OPERATING SYSTEM: Microsoft Windows 95  
SOFTWARE: Wordperfect for windows 6.0,  
SOFTWARE: ASCII (DOS) TEXT format  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/324,709A  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
ATTORNEY/AGENT INFORMATION:  
NAME: David J. Weitz  
REGISTRATION NUMBER: 38,362  
REFERENCE/DOCKET NUMBER: PELM-744  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 493-9300  
TELEFAX: (650) 493-6811  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 793 nucleotides  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-018-595B-1

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 493-9300  
TELEFAX: (650) 493-6811  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 500 nucleotides  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-09-324-709A-3

Query Match 48.3%; Score 33.8; DB 3; Length 500;  
Best Local Similarity 84.4%; Pred. No. 0.0013;  
Matches 38; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 GATTTGTTGGCTGCTCTGGAGAGAGCTTTCAGTATGCTGT 45  
Db 80 GATTTATTTGCTGCTCTGAGAGAGAGCTTTGCGATGCTCT 124

RESULT 4  
US-09-018-595B-1  
Sequence 1, Application US/09018595B  
Patent No. 5962233  
GENERAL INFORMATION:  
APPLICANT: Perkin-Elmer Corporation,  
APPLICANT: Applied Biosystems Division  
TITLE OF INVENTION: DETERMINATION OF GENOTYPE OF  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David J. Weitz,  
ADDRESSEE: Wilson Sonsini Goodrich & Rosati  
STREET: 650 Page Mill Road  
CITY: Palo Alto  
STATE: California  
COUNTRY: USA  
ZIP: 94304-1050  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch diskette  
COMPUTER: IBM compatible  
OPERATING SYSTEM: Microsoft Windows 95  
SOFTWARE: Wordperfect for windows 6.0,  
SOFTWARE: ASCII (DOS) TEXT format  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/018,595B  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
ATTORNEY/AGENT INFORMATION:  
NAME: David J. Weitz  
REGISTRATION NUMBER: 38,362  
REFERENCE/DOCKET NUMBER: PELM-744  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 493-9300  
TELEFAX: (650) 493-6811  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 793 nucleotides  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-018-595B-1

Query Match 48.3%; Score 33.8; DB 2; Length 793;  
Best Local Similarity 84.4%; Pred. No. 0.0015;  
Matches 38; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 GATTTGTTGGCTGCTCTGGAGAGAGCTTTCAGTATGCTGT 45  
Db 80 GATTTATTTGCTGCTCTGAGAGAGAGCTTTGCGATGCTCT 124





```

RESULT 8
US-09-205-258-13/c
: Sequence 13, Application US/09205258
: Patent No. 6525174
: GENERAL INFORMATION:
: APPLICANT: Young et al.
: TITLE OF INVENTION: 207 Human Secreted Proteins
: FILE REFERENCE: P2007P1
: CURRENT APPLICATION NUMBER: US/09/205,258
: CURRENT FILING DATE: 1998-12-04
: EARLIER APPLICATION NUMBER: PCT/US98/11422
: EARLIER FILING DATE: 1998-06-04
: EARLIER APPLICATION NUMBER: 60/048,885
: EARLIER FILING DATE: 1997-06-06
: EARLIER APPLICATION NUMBER: 60/049,375
: EARLIER FILING DATE: 1997-06-06
: EARLIER APPLICATION NUMBER: 60/048,881
: EARLIER FILING DATE: 1997-06-06
: EARLIER APPLICATION NUMBER: 60/048,880
: EARLIER FILING DATE: 1997-06-06
: EARLIER APPLICATION NUMBER: 60/048,896
: EARLIER FILING DATE: 1997-06-06
: EARLIER APPLICATION NUMBER: 60/049,020
: EARLIER FILING DATE: 1997-06-06
: EARLIER APPLICATION NUMBER: 60/048,876
: EARLIER FILING DATE: 1997-06-06
: EARLIER APPLICATION NUMBER: 60/048,895
: EARLIER FILING DATE: 1997-06-06
: EARLIER APPLICATION NUMBER: 60/048,884
: EARLIER FILING DATE: 1997-06-06
: EARLIER APPLICATION NUMBER: 60/048,894
: EARLIER FILING DATE: 1997-06-06
: EARLIER APPLICATION NUMBER: 60/048,971
: EARLIER FILING DATE: 1997-06-06
: EARLIER APPLICATION NUMBER: 60/048,964
: EARLIER FILING DATE: 1997-06-06
: EARLIER APPLICATION NUMBER: 60/048,882
: EARLIER FILING DATE: 1997-06-06
: EARLIER APPLICATION NUMBER: 60/048,899
: EARLIER FILING DATE: 1997-06-06
: EARLIER APPLICATION NUMBER: 60/048,893
: EARLIER FILING DATE: 1997-06-06
: EARLIER APPLICATION NUMBER: 60/048,900
: EARLIER FILING DATE: 1997-06-06
: EARLIER APPLICATION NUMBER: 60/048,901
: EARLIER FILING DATE: 1997-06-06
: EARLIER APPLICATION NUMBER: 60/048,892
: EARLIER FILING DATE: 1997-06-06
: EARLIER APPLICATION NUMBER: 60/048,915
: EARLIER FILING DATE: 1997-06-06
: EARLIER APPLICATION NUMBER: 60/049,019
: EARLIER FILING DATE: 1997-06-06

```

Query Match	37.7%	Score 26.4	DB 4	Length 941
Best Local Similarity	69.2%	Pred. No. 1.2		
Matches	36	Conservative	0	Mismatches 16; Indels 0; Gaps 0
Qy	17	CTCTGGGAGCAGCCTTCAGTATCCCTGTGAGTAAATTTCCAAATTTCCAAAT	68	
Db	845	CTACTGGACACGCTTAAATGACCTTCAGCTTAAATTTTCGTTTCCAAAT	794	

```

RESULT 9
US-09-205-258-218/c
Sequence 218, Application US/09205258
Patent No. 6525174
GENERAL INFORMATION:
APPLICANT: Young et al.
TITLE OR INVENTION: 207 Human Secreted Proteins
FILE REFERENCE: P2007P1
CURRENT APPLICATION NUMBER: US/09/205,258
EARLIER FILING DATE: 1998-12-04
EARLIER APPLICATION NUMBER: PCT/US98/11422
EARLIER FILING DATE: 1998-06-04
EARLIER APPLICATION NUMBER: 60/048,885
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/049,375
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,881
EARLIER FILING DATE: 1997-06-06

```

EARLIER APPLICATION NUMBER: 60/048,880  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,896  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/049,020  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,876  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,895  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,884  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,894  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,971  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,964  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,882  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,899  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,893  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,900  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,901  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,892  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,915  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/049,019  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,970  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,972  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,916  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/049,373  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,875  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/049,374  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,917  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,949  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,974  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,883  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,897  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,898  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,962  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,963  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,877  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,878  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/070,923  
EARLIER FILING DATE: 1997-12-18  
EARLIER APPLICATION NUMBER: 60/092,921  
EARLIER FILING DATE: 1998-07-15  
EARLIER APPLICATION NUMBER: 60/094,657  
EARLIER FILING DATE: 1998-07-30  
NUMBER OF SEQ ID NOS: 1227

SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 218  
LENGTH: 941  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-205-258-218

Query Match 37.7%; Score 26.4; DB 4; Length 941;  
Best Local Similarity 69.2%; Pred. No. 1.2;  
Matches 36; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

17 CTCCTGGAGACGCTTCAAGTATCCGCGTGAATAATTCATTTCCAT 68  
DB 845 CTACTCGACACGCTTAATGCACTTCACTTAATTTTGTTCAT 794

RESULT 10  
US-09-513-999C-32153  
Sequence 32153; Application US/09513999C  
Patent No. 6783961  
GENERAL INFORMATION:  
APPLICANT: Dumas Milne Edwards, J.B.  
APPLICANT: Duclert, A.  
APPLICANT: Giordano, J.Y.  
TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.  
PATENT REFERENCE: 59.US2.REG  
CURRENT APPLICATION NUMBER: US/09/513,999C  
CURRENT FILING DATE: 2000-02-24  
PRIOR APPLICATION NUMBER: US 60/122,487  
PRIOR FILING DATE: 1999-02-26  
NUMBER OF SEQ ID NOS: 36681  
SOFTWARE: Patent.pm  
SEQ ID NO 32153  
LENGTH: 292  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-513-999C-32153

Query Match 36.9%; Score 25.8; DB 4; Length 292;  
Best Local Similarity 60.9%; Pred. No. 1.3;  
Matches 42; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 2 ATTTGTTGCTCTGCTGAGACGCTTGAATGCTGTGATAATTTCCAT 61  
DB 100 ATTTTATTTCTTATTTAGGAGGTGCTACAAATGTTGTCACTTAATTTCAAGT 159  
QY 62 TTCCAATTT 70  
DB 160 TTCTGTTT 168

RESULT 11  
US-09-328-352-134  
Sequence 134; Application US/09328352  
Patent No. 6562958  
GENERAL INFORMATION:  
APPLICANT: Gary L. Breton et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
FILE REFERENCE: GTC99-03PA  
CURRENT APPLICATION NUMBER: US/09/328,352  
CURRENT FILING DATE: 1999-06-04  
NUMBER OF SEQ ID NOS: 8252  
SEQ ID NO 134  
LENGTH: 2871  
TYPE: DNA  
ORGANISM: Acinetobacter baumannii  
US-09-328-352-134

Query Match 35.7%; Score 25; DB 4; Length 2871;  
Best Local Similarity 64.9%; Pred. No. 6.1;  
Matches 37; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 14 TGGCTCTGGGAGACGCTTCAGTANGCCTGTGAGTAAATTTCCAAATTTCCAAATTT 70  
DB 737 TGGCGCAAAATAGACCCGATTATGAGCCAGATATTTATTTCCAAATTCAAATTT 793

## RESULT 12

US-09-270-767-9938/C  
Sequence 9938, Application US/09270767  
Patent No. 6703491  
GENERAL INFORMATION:  
APPLICANT: Homburger et al.  
TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster  
FILE REFERENCE: File Reference: 7326-094  
CURRENT APPLICATION NUMBER: US/09/270,767  
CURRENT FILING DATE: 1999-03-17  
NUMBER OF SEQ ID NOS: 62517  
SOFTWARE: Patentln Ver. 2.0  
SEQ ID NO 9938  
LENGTH: 756  
TYPE: DNA  
ORGANISM: Drosophila melanogaster  
US-09-270-767-9938

Query Match 35.4%; Score 24.8; DB 4; Length 756;  
Best Local Similarity 63.3%; Pred. No. 4.4;  
Matches 38; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 2 ATTTTGTTCCTGCTCCTCTGGAGACGCTTCAGTATGCTGTGAGTAAATTTCCAAATTT 61  
DB 181 ATTTGAAATCATCTACTGTGTGACGAAAGAGTGTGCTTCAGTAAATTTCCAAATTT 122

## RESULT 13

US-09-270-767-13233/C  
Sequence 13233, Application US/09270767  
Patent No. 6703491  
GENERAL INFORMATION:  
APPLICANT: Homburger et al.  
TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster  
FILE REFERENCE: File Reference: 7326-094  
CURRENT APPLICATION NUMBER: US/09/270,767  
CURRENT FILING DATE: 1999-03-17  
NUMBER OF SEQ ID NOS: 62517  
SOFTWARE: Patentln Ver. 2.0  
SEQ ID NO 13233  
LENGTH: 1863  
TYPE: DNA  
ORGANISM: Drosophila melanogaster  
US-09-270-767-13233

Query Match 35.4%; Score 24.8; DB 4; Length 1863;  
Best Local Similarity 60.3%; Pred. No. 6.2;  
Matches 41; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 3 TTTTGTTCCTGCTCCTCTGGAGACGCTTCAGTATGCTGTGAGTAAATTTCCAAATTT 62  
DB 302 TTTTGTTCCTTTTACCCCGCTGCGGTCTGCGGCTTGCTTTTGTGTTTAAATTT 243

QY 63 TCCAAATTT 70  
DB 242 TCTCTCTT 235

## RESULT 14

US-08-896-164-80/C  
Sequence 80, Application US/08896164  
Patent No. 6218521  
GENERAL INFORMATION:  
APPLICANT: OBATA, YUICHI  
TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULES ASSOCIATED  
TITLE OF INVENTION: WITH GASTRIC CANCER AND METHODS FOR  
TITLE OF INVENTION: DIAGNOSING AND TREATING GASTRIC CANCER

NUMBER OF SEQUENCES: 87  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Felle & Lynch  
STREET: 805 Third Avenue  
CITY: New York City  
STATE: New York  
COUNTRY: USA  
ZIP: 10022

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.44mb  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: PC-DOS  
SOFTWARE: WordPerfect  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/896,164  
FILING DATE: July 17, 1997  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 6218521man D. Hanson  
REGISTRATION NUMBER: 30,946  
REFERENCE/DOCKET NUMBER: LUD 5499 - JEL/NDH/SLH  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 688-9200  
TELEFAX: (212) 838-3884  
INFORMATION FOR SEQ ID NO: 80:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 592 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-896-164-80

Query Match 34.3%; Score 24; DB 3; Length 592;  
Best Local Similarity 59.1%; Pred. No. 8.2;  
Matches 39; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 5 TTGTTGCTGCTCCTCTCTGGAGACGCTTCAGTATGCTGTGAGTAAATTTCCAAATTT 64  
DB 562 TTGTTGCTGCTCCTCTCTGGAGATCTGATGCTGTGAGTAAATTTCCAAATTT 503

QY 65 CAATTT 70  
DB 502 TATTTT 497

## RESULT 15

US-08-266-080B-1/C  
Sequence 1, Application US/08266080B  
Patent No. 5606031  
GENERAL INFORMATION:  
APPLICANT: Jack Lille  
APPLICANT: Tadahiko Kohno  
APPLICANT: Duane Bonam  
APPLICANT: Mary S. Rosendahl  
TITLE OF INVENTION: Production of Biologically Active  
TITLE OF INVENTION: Recombinant Neurotrophic Protein  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Swanson & Bratschun, L.L.C.  
STREET: 8400 E. Prentice Avenue, Suite 200  
CITY: Englewood  
STATE: Colorado  
COUNTRY: USA  
ZIP: 80111  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 5.25 inch, 360 Kb storage  
COMPUTER: IBM compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: WordPerfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/266,080B  
FILING DATE: 27-JUNE-1994  
PRIOR APPLICATION DATA:

```

1  APPLICATION NUMBER: 08/240,122
2  FILING DATE: 09-MAY-1994
3  PRIOR APPLICATION DATA:
4  APPLICATION NUMBER: 08/087,912
5  FILING DATE: 06-JULY-1993
6  PRIOR APPLICATION DATA:
7  APPLICATION NUMBER: 07/680,681
8  FILING DATE: 04-APRIL-1991
9  PRIOR APPLICATION DATA:
10 APPLICATION NUMBER: 07/594,126
11 FILING DATE: 09-OCT-1990
12 PRIOR APPLICATION DATA:
13 APPLICATION NUMBER: 07/547,750
14 FILING DATE: 02-JULY-1990
15 PRIOR APPLICATION DATA:
16 APPLICATION NUMBER: 07/505,441
17 FILING DATE: 06-APRIL-1990
18 ATTORNEY/AGENT INFORMATION:
19 NAME: Barry J. Swanson
20 REGISTRATION NUMBER: 33,215
21 REFERENCE/DOCKET NUMBER: SYNE200CS
22 TELECOMMUNICATION INFORMATION:
23 TELEPHONE: (303) 793-3333
24 TELEFAX: (303) 793-4333
25 INFORMATION FOR SEQ ID NO: 1:
26 SEQUENCE CHARACTERISTICS:
27 LENGTH: 742 nucleotides
28 TYPE: nucleic acid
29 STRANDEDNESS: single
30 TOPOLOGY: linear
31 FEATURE:
32 NAME/KEY: nucleic acid sequence for human BDNF
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
9
```

This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 8, 2005, 16:37:21 ; Search time 94.3248 Seconds  
(without alignments)  
4264.119 Million cell updates/sec

Title: US-10-754-437-7

Perfect score: 70

Sequence: 1 gatcttgctgcctgcctcc.....aaattccaatccaatc 70

Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4300275 seqs, 2872944193 residues

Total number of hits satisfying chosen parameters: 8600550

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing:

Maximum Match 0%  
Listing first 45 summaries

Database :

Published Applications NA:\*

- 1: /cgn2\_6/prodata/2/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/prodata/2/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/prodata/2/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/prodata/2/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/prodata/2/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/prodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/prodata/2/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/prodata/2/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/prodata/2/pubpna/US09A\_PUBCOMB.seq:\*
- 10: /cgn2\_6/prodata/2/pubpna/US09B\_PUBCOMB.seq:\*
- 11: /cgn2\_6/prodata/2/pubpna/US09C\_PUBCOMB.seq:\*
- 12: /cgn2\_6/prodata/2/pubpna/US09\_NEW\_PUB.seq:\*
- 13: /cgn2\_6/prodata/2/pubpna/US10\_PUBCOMB.seq:\*
- 14: /cgn2\_6/prodata/2/pubpna/US10B\_PUBCOMB.seq:\*
- 15: /cgn2\_6/prodata/2/pubpna/US10C\_PUBCOMB.seq:\*
- 16: /cgn2\_6/prodata/2/pubpna/US10D\_PUBCOMB.seq:\*
- 17: /cgn2\_6/prodata/2/pubpna/US10E\_PUBCOMB.seq:\*
- 18: /cgn2\_6/prodata/2/pubpna/US10\_NEW\_PUB.seq:\*
- 19: /cgn2\_6/prodata/2/pubpna/US11\_NEW\_PUB.seq:\*
- 20: /cgn2\_6/prodata/2/pubpna/US60\_NEW\_PUB.seq:\*
- 21: /cgn2\_6/prodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	43.6	62.3	600	US-10-027-632-287049	Sequence 287049,
C 2	43.6	62.3	600	US-10-027-632-287050	Sequence 287050,
C 3	43.6	62.3	600	US-10-027-632-287049	Sequence 287049,
C 4	43.6	62.3	600	US-10-027-632-287050	Sequence 287050,
C 5	43.6	62.3	600	US-10-027-632-287050	Sequence 287050,
C 6	43.6	62.3	600	US-10-027-632-287050	Sequence 287050,
C 7	43.6	62.3	600	US-10-027-632-287050	Sequence 287050,
C 8	43.6	62.3	600	US-10-027-632-287050	Sequence 287050,
C 9	43.6	62.3	600	US-10-027-632-287050	Sequence 287050,
C 10	43.6	62.3	600	US-10-027-632-287050	Sequence 287050,
C 11	43.6	62.3	600	US-10-027-632-287050	Sequence 287050,
C 12	43.6	62.3	600	US-10-027-632-287050	Sequence 287050,

13	32.2	46.0	802	13	US-10-104-774-2	Sequence 2, Appl1
14	32.2	46.0	802	16	US-10-455-150-2	Sequence 2, Appl1
15	32.2	46.0	802	18	US-10-723-860-5797	Sequence 5797, Ap
16	28.6	40.9	277	18	US-10-856-499-38	Sequence 38, Appl
17	27.8	39.7	1185	15	US-10-369-493-34889	Sequence 34889, A
18	27.2	38.9	129	17	US-10-437-963-53858	Sequence 53858, A
19	26.6	38.0	653122	13	US-10-087-192-226	Sequence 226, App
20	26.4	37.7	423	9	US-09-864-761-11109	Sequence 11109, A
21	26.4	37.7	567	9	US-09-864-761-16464	Sequence 16464, A
22	26.4	37.7	941	10	US-09-933-767-13	Sequence 13, Appl
23	26.4	37.7	941	10	US-09-933-767-218	Sequence 218, Appl
24	26.4	37.7	941	14	US-10-004-860-13	Sequence 13, Appl
25	26.4	37.7	941	14	US-10-004-860-218	Sequence 218, Appl
26	26.4	37.7	941	14	US-10-023-282-13	Sequence 13, Appl
27	26.4	37.7	941	14	US-10-023-282-218	Sequence 218, Appl
28	26.4	37.7	1658	13	US-10-027-632-264601	Sequence 264601,
29	26.4	37.7	1658	13	US-10-027-632-264602	Sequence 264602,
30	26.4	37.7	1658	15	US-10-027-632-264601	Sequence 264601,
31	26.4	37.7	1658	15	US-10-027-632-264602	Sequence 264602,
32	25.8	36.9	2079	13	US-10-027-632-97628	Sequence 97628, A
33	25.8	36.9	2079	13	US-10-027-632-100426	Sequence 100426,
34	25.8	36.9	2079	13	US-10-027-632-100427	Sequence 100427,
35	25.8	36.9	2079	15	US-10-027-632-97628	Sequence 97628, A
36	25.8	36.9	2079	15	US-10-027-632-100426	Sequence 100426,
37	25.8	36.9	2079	15	US-10-027-632-100427	Sequence 100427,
38	25.8	36.9	3537	15	US-10-369-493-46684	Sequence 46684, A
39	25.8	36.9	66973	13	US-10-087-192-574	Sequence 574, Appl
40	25.8	36.9	74665	18	US-10-719-993-6854	Sequence 6854, Ap
41	25.4	36.3	410	16	US-10-424-599-16385	Sequence 16385, A
42	25.2	36.0	1650	14	US-10-198-846-11735	Sequence 11735, A
43	25.2	36.0	48436	9	US-09-927-602-38	Sequence 38, Appl
44	25.2	36.0	96588	11	US-09-997-722-292	Sequence 292, App
45	25.2	35.7	661	13	US-10-027-632-273400	Sequence 273400,

#### ALIGNMENTS

RESULT 1  
US-10-027-632-287049/c  
; Sequence 287049, Application US/10027632  
; Publication No. US20020198371A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; TITLE OF INVENTION: Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO: 287049  
; LENGTH: 600  
; TYPE: DNA  
; ORGANISM: Human  
; US-10-027-632-287049

Query Match 62.3%; Score 43.6; DB 13; Length 600;  
Best Local Similarity 84.5%; Pred. No. 3e-06;

Matches 49; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

OY 1 GATTTGTTTGGCTTCTGCGAGAGAGCTTACATGCTGTGAGTAAATTTCC 58  
DB 245 GATTTATTTGCTGCTGCTGCGAGAGAGCTTTGGCATGCTGTGAGTAAACACCC 188

## RESULT 2

US-10-027-632-287050/c  
Sequence 287050, Application US/10027632  
Publication No. US2002019837A1  
GENERAL INFORMATION:  
APPLICANT: Wang, David G.  
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
FILE REFERENCE: 108827.129  
CURRENT FILING DATE: 2002-04-30  
PRIOR FILING DATE: 2000-07-12  
PRIOR APPLICATION NUMBER: US 60/218,006  
PRIOR FILING DATE: 2000-07-12  
PRIOR APPLICATION NUMBER: US 60/198,676  
PRIOR FILING DATE: 2000-04-20  
PRIOR APPLICATION NUMBER: US 60/193,483  
PRIOR FILING DATE: 2000-03-29  
PRIOR APPLICATION NUMBER: US 60/185,218  
PRIOR FILING DATE: 2000-02-24  
PRIOR APPLICATION NUMBER: US 60/167,363  
PRIOR FILING DATE: 1999-11-23  
PRIOR APPLICATION NUMBER: US 60/156,358  
PRIOR FILING DATE: 1999-09-28  
PRIOR APPLICATION NUMBER: US 60/146,002  
PRIOR FILING DATE: 1999-08-09  
NUMBER OF SEQ ID NOS: 325720  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 287050  
LENGTH: 600  
TYPE: DNA  
ORGANISM: Human  
US-10-027-632-287050

Query Match 62.3%; Score 43.6; DB 13; Length 600;

Best Local Similarity 84.5%; Pred. No. 3e-06; Mismatches 9; Indels 0; Gaps 0;

OY 1 GATTTGTTTGGCTTCTGCGAGAGAGCTTACATGCTGTGAGTAAATTTCC 58  
DB 245 GATTTATTTGCTGCTGCTGCGAGAGAGCTTTGGCATGCTGTGAGTAAACACCC 188

## RESULT 3

US-10-027-632-287049/c  
Sequence 287049, Application US/10027632  
Publication No. US20030204075A9  
GENERAL INFORMATION:  
APPLICANT: Wang, David G.  
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
FILE REFERENCE: 108827.129  
CURRENT FILING DATE: 2002-04-30  
PRIOR FILING DATE: 2002-04-30  
PRIOR APPLICATION NUMBER: US 60/218,006  
PRIOR FILING DATE: 2000-07-12  
PRIOR APPLICATION NUMBER: US 60/198,676  
PRIOR FILING DATE: 2000-04-20  
PRIOR APPLICATION NUMBER: US 60/193,483  
PRIOR FILING DATE: 2000-03-29  
PRIOR APPLICATION NUMBER: US 60/185,218  
PRIOR FILING DATE: 2000-02-24  
PRIOR APPLICATION NUMBER: US 60/167,363  
PRIOR FILING DATE: 1999-11-23  
PRIOR APPLICATION NUMBER: US 60/156,358  
PRIOR FILING DATE: 1999-09-28  
PRIOR APPLICATION NUMBER: US 60/146,002

PRIOR FILING DATE: 1999-08-09

NUMBER OF SEQ ID NOS: 325720

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 287049

LENGTH: 600

TYPE: DNA

ORGANISM: Human

US-10-027-632-287049

OY 1 GATTTGTTTGGCTTCTGCGAGAGAGCTTACATGCTGTGAGTAAATTTCC 58  
DB 245 GATTTATTTGCTGCTGCTGCGAGAGAGCTTTGGCATGCTGTGAGTAAACACCC 188

## RESULT 4

US-10-027-632-287050/c  
Sequence 287050, Application US/10027632  
Publication No. US20030204075A9  
GENERAL INFORMATION:  
APPLICANT: Wang, David G.  
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
FILE REFERENCE: 108827.129  
CURRENT FILING DATE: 2002-04-30  
PRIOR FILING DATE: 2000-07-12  
PRIOR APPLICATION NUMBER: US 60/218,006  
PRIOR FILING DATE: 2000-07-12  
PRIOR APPLICATION NUMBER: US 60/198,676  
PRIOR FILING DATE: 2000-04-20  
PRIOR APPLICATION NUMBER: US 60/193,483  
PRIOR FILING DATE: 2000-03-29  
PRIOR APPLICATION NUMBER: US 60/185,218  
PRIOR FILING DATE: 2000-02-24  
PRIOR APPLICATION NUMBER: US 60/167,363  
PRIOR FILING DATE: 1999-11-23  
PRIOR APPLICATION NUMBER: US 60/156,358  
PRIOR FILING DATE: 1999-09-28  
PRIOR APPLICATION NUMBER: US 60/146,002  
PRIOR FILING DATE: 1999-08-09  
NUMBER OF SEQ ID NOS: 325720  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 287050  
LENGTH: 600  
TYPE: DNA  
ORGANISM: Human  
US-10-027-632-287050

Query Match 62.3%; Score 43.6; DB 15; Length 600;

Best Local Similarity 84.5%; Pred. No. 3e-06; Mismatches 9; Indels 0; Gaps 0;

OY 1 GATTTGTTTGGCTTCTGCGAGAGAGCTTACATGCTGTGAGTAAATTTCC 58  
DB 245 GATTTATTTGCTGCTGCTGCGAGAGAGCTTTGGCATGCTGTGAGTAAACACCC 188

## RESULT 5

US-10-723-860-1311  
Sequence 1311, Application US/10723860  
Publication No. US20040253606A1  
GENERAL INFORMATION:  
APPLICANT: Aziz, Nataasha  
APPLICANT: Gineburg, Wendy M.  
APPLICANT: Zlotnik, Albert  
TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &  
FILE REFERENCE: 05882.0193.NPUS01  
CURRENT FILING DATE: 2003-11-26



PRIOR APPLICATION NUMBER: 60/429,739  
PRIOR FILING DATE: 2002-11-26  
NUMBER OF SEQ ID NOS: 8393  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 1311  
LENGTH: 259202  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-723-860-1311

Query Match 62.3%; Score 43.6; DB 18; Length 259202;  
Best Local Similarity 84.5%; Pred. No. 1.8e-05;  
Matches 49; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 1 GATTTGTTGCTGCTCTGAGAGAGCCTTCAGTATGCTGTGATTAATTTCC 58  
Db 69884 GATTTATTTGCTGCTCTGAGAGAGCCTTTGCTATGCTGTGATTAACACCC 69941

## RESULT 6

US-10-723-860-6937/C  
Sequence 6937, Application US/10723860  
Publication No. US20040253606A1  
GENERAL INFORMATION:  
APPLICANT: Aziz, Nafasha  
APPLICANT: Gansburg, Wendy M.  
APPLICANT: Zlotnick, Albert  
TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &  
FILE REFERENCE: 05882.0193.NPUS01  
CURRENT APPLICATION NUMBER: US/10/723,860  
PRIOR FILING DATE: 2003-11-26  
PRIOR APPLICATION NUMBER: 60/429,739  
NUMBER OF SEQ ID NOS: 8393  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 6937  
LENGTH: 3037  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: (1804)..(1823)  
OTHER INFORMATION: n 1a a, c, g, or t  
US-10-723-860-6937

Query Match 50.9%; Score 35.6; DB 18; Length 3037;  
Best Local Similarity 71.2%; Pred. No. 0.0053;  
Matches 47; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

Qy 4 TTTGTTGCTGCTCTGAGAGAGCCTTCAGTATGCTGTGATTAATTTCCATTT 63  
Db 913 TTTCTGCTGCTGCTCTGAGAGAGCCTTAATTCAGTTTCTGTTATTAACATTTACCTTT 854

Qy 64 CCAATT 69  
Db 853 CTCCTT 848

## RESULT 7

US-09-908-975-1137  
Sequence 1137, Application US/09908975  
Publication No. US20030165843A1  
GENERAL INFORMATION:  
APPLICANT: SHOSHAN, Avi  
APPLICANT: WASSERMAN, Alon  
APPLICANT: MINTZ, Eli  
APPLICANT: MINTZ, Liat  
APPLICANT: FAIGER, Simchon  
TITLE OF INVENTION: OLIGONUCLEOTIDE LIBRARY FOR DETECTING RNA TRANSCRIPTS AND SPLICE  
FILE REFERENCE: 36688-0005  
CURRENT APPLICATION NUMBER: US/09/908,975

CURRENT FILING DATE: 2001-07-20  
PRIOR APPLICATION NUMBER: US 60/287,724  
PRIOR FILING DATE: 2001-05-02  
PRIOR APPLICATION NUMBER: US 60/221,607  
PRIOR FILING DATE: 2000-07-28  
NUMBER OF SEQ ID NOS: 32337  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 1137  
LENGTH: 65  
TYPE: DNA  
ORGANISM: Rattus norvegicus  
US-09-908-975-1137

Query Match 48.6%; Score 34; DB 10; Length 65;  
Best Local Similarity 88.1%; Pred. No. 0.0063;  
Matches 37; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 GATTTGTTGCTGCTCTGAGAGAGCCTTCAGTATGCC 42  
Db 6 GATCTTGTGCTGCTCTGAGAGAGCCTTTGCTATGCC 47

## RESULT 8

US-10-104-774-3  
Sequence 3, Application US/10104774  
Publication No. US20020164630A1  
GENERAL INFORMATION:  
APPLICANT: Perkin-Elmer Corporation, Applied Biosystems Division  
TITLE OF INVENTION: DETERMINATION OF GENOTYPE OF AMPLIFICATION PRODUCTS AT MULTIPLE  
FILE REFERENCE: 16842-782  
CURRENT APPLICATION NUMBER: US/10/104,774  
PRIOR FILING DATE: 2002-03-21  
PRIOR APPLICATION NUMBER: US 09/018,595  
PRIOR FILING DATE: 1998-02-04  
NUMBER OF SEQ ID NOS: 11  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 3  
LENGTH: 500  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-104-774-3

Query Match 48.3%; Score 33.8; DB 13; Length 500;  
Best Local Similarity 84.4%; Pred. No. 0.015;  
Matches 38; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 GATTTGTTGCTGCTCTGAGAGAGCCTTCAGTATGCCGT 45  
Db 80 GATTTATTTGCTGCTCTGAGAGAGCCTTTGCCATGCCCTT 124

## RESULT 9

US-10-455-150-3  
Sequence 3, Application US/10455150  
Publication No. US20040053302A1  
GENERAL INFORMATION:  
APPLICANT: Perkin-Elmer Corporation,  
TITLE OF INVENTION: Applied Biosystems Division  
FLUORESCENCE GENOTYPING AT MULTIPLE ALLELIC  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSER: David J. Weitz,  
Wilson Sonsini Goodrich & Rosati  
STREET: 650 Page Mill Road  
CITY: Palo Alto  
STATE: California  
COUNTRY: USA  
ZIP: 94304-1050  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch diskette  
COMPUTER: IBM compatible

```
OPERATING SYSTEM: Microsoft Windows 95
SOFTWARE: Wordperfect for windows 6.0,
ASCII (DOS) TEXT format
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/455,150
FILING DATE: 04-Jun-2003
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/324,709A
FILING DATE: 03-June-1999
ATTORNEY/AGENT INFORMATION:
NAME: David J. Weitz
REGISTRATION NUMBER: 38,362
REFERENCE/DOCKET NUMBER: 16842-758
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 493-9300
TELEFAX: (650) 493-6811
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 500 nucleotides
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-10-455-150-3

Query Match      48.3%; Score 33.8; DB 16; Length 500;
Best Local Similarity 84.4%; Pred. No. 0.015; Indels 0; Gaps 0;
Matches 38; Conservative 0; Mismatches 7;

Qy      1 GATTGTGTTGCTGCTCTGGAGAGACGCTTCAGTATGCTGT 45
Db      80 GATTATTGCTGCTGCTCTGGAGAGACGCTTTGCCATGCTCT 124

RESULT 10
US-10-104-774-1
Sequence 1, Application US/10104774
Publication No. US20020164630A1
GENERAL INFORMATION:
APPLICANT: Perkin-Elmer Corporation, Applied Biosystems Division
TITLE OF INVENTION: DETERMINATION OF GENOTYPE OF AMPLIFICATION PRODUCTS AT MULTIPLE A
FILE REFERENCE: 16842-782
CURRENT APPLICATION NUMBER: US/10/104,774
CURRENT FILING DATE: 2002-03-21
PRIOR APPLICATION NUMBER: US 09/018,595
PRIOR FILING DATE: 1998-02-04
NUMBER OF SEQ ID NOS: 11
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1
LENGTH: 793
TYPE: DNA
ORGANISM: Homo sapiens
US-10-104-774-1

Query Match      48.3%; Score 33.8; DB 13; Length 793;
Best Local Similarity 84.4%; Pred. No. 0.017; Indels 7; Gaps 0;
Matches 38; Conservative 0; Mismatches 7;

Qy      1 GATTGTGTTGCTGCTCTGGAGAGACGCTTCAGTATGCTGT 45
Db      80 GATTATTGCTGCTGCTCTGGAGAGACGCTTTGCCATGCTCT 124

RESULT 11
US-10-455-150-1
Sequence 1, Application US/10455150
Publication No. US20040053302A1
GENERAL INFORMATION:
APPLICANT: Perkin-Elmer Corporation,
Applied Biosystems Division
TITLE OF INVENTION: FLUORESCENCE GENOTYPING AT MULTIPLE ALLELIC
```

```
SITES
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSER: David J. Weitz,
Wilson Sonaini Goodrich & Rosati
STREET: 650 Page Mill Road
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1050
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: Microsoft Windows 95
SOFTWARE: Wordperfect for windows 6.0,
ASCII (DOS) TEXT format
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/455,150
FILING DATE: 04-Jun-2003
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/324,709A
FILING DATE: 03-June-1999
ATTORNEY/AGENT INFORMATION:
NAME: David J. Weitz
REGISTRATION NUMBER: 38,362
REFERENCE/DOCKET NUMBER: 16842-758
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 493-9300
TELEFAX: (650) 493-6811
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 793 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-10-455-150-1

Query Match      48.3%; Score 33.8; DB 16; Length 793;
Best Local Similarity 84.4%; Pred. No. 0.017; Indels 7; Gaps 0;
Matches 38; Conservative 0; Mismatches 7;

Qy      1 GATTGTGTTGCTGCTCTGGAGAGACGCTTCAGTATGCTGT 45
Db      80 GATTATTGCTGCTGCTCTGGAGAGACGCTTTGCCATGCTCT 124

RESULT 12
US-10-755-889-605
Sequence 605, Application US/10755889
Publication No. US20040171823A1
GENERAL INFORMATION:
APPLICANT: Bristol-Myers Squibb Company
TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ASSOCIATED WITH THE NF-KB
FILE REFERENCE: D0284 NP
CURRENT APPLICATION NUMBER: US/10/755,889
CURRENT FILING DATE: 2004-01-13
PRIOR APPLICATION NUMBER: U.S. 60/440,068
PRIOR FILING DATE: 2003-01-14
PRIOR APPLICATION NUMBER: U.S. 60/469,757
PRIOR FILING DATE: 2003-05-12
NUMBER OF SEQ ID NOS: 823
SOFTWARE: PatentIn version 3.2
SEQ ID NO 605
LENGTH: 793
TYPE: DNA
ORGANISM: Homo sapiens
US-10-755-889-605

Query Match      48.3%; Score 33.8; DB 17; Length 793;
Best Local Similarity 84.4%; Pred. No. 0.017; Indels 7; Gaps 0;
```



**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 8, 2005, 12:58:25 ; Search time 833.766 Seconds  
(without alignments)  
3059.349 Million cell updates/sec

Title: US-10-754-437-7

Perfect score: 70  
Sequence: 1 gatttcgttcgtcgtcgtcc.....aaattccaatccattc 70

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 32822875 seqs, 1821965908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

EST:\*  
1: gb\_esc1:\*  
2: gb\_esc2:\*  
3: gb\_esc3:\*  
4: gb\_esc4:\*  
5: gb\_esc5:\*  
6: gb\_esc6:\*  
7: gb\_esc7:\*  
8: gb\_esc8:\*  
9: gb\_esc9:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysts of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	38.6	55.1	439	6	CB473334 sn68_E03.
2	35.6	50.9	303	7	W33932 md54d02.r1
3	35.6	50.9	430	7	W40649 mc43h12.r1
4	35.6	50.9	441	1	A1798381 tr34c02.x
5	35.6	50.9	444	2	BE220076 hv6h03.x
6	35.6	50.9	448	1	A1338491 q997h05.x
7	35.6	50.9	536	7	W36345 mb72c12.r1
8	35.6	50.9	561	5	BX470683 DKZP686N
9	35.6	50.9	585	9	AY149441 Mus muscu
10	35.6	50.9	586	2	BF001849 7994c09.x
11	35.6	50.9	588	4	BF448291 nad17f04.
12	35.6	50.9	620	2	BG622884 602647718
13	35.6	50.9	645	6	CD773419 AGENCOURT
14	35.6	50.9	707	6	CB056709 NISC_j119
15	35.6	50.9	707	7	CN360063 170005830
16	35.6	50.9	825	3	AK029358 Mus muscu
17	35.6	50.9	825	1	AU118731 AU118731
18	35.6	50.9	853	6	CB588212 AGENCOURT
19	35.6	50.9	859	6	CB588525 AGENCOURT
20	35.6	50.9	865	6	CB589177 AGENCOURT
21	35.6	50.9	886	6	CB587051 AGENCOURT
22	35.6	50.9	889	6	CB589251 AGENCOURT
23	35.6	50.9	891	6	CB574837 AGENCOURT
24	35.6	50.9	913	6	CB590451 AGENCOURT

25	35.6	50.9	919	6	CB587332 AGENCOURT
26	35.6	50.9	928	6	CB590111 AGENCOURT
27	35.6	50.9	945	2	BB614068 BB614068
28	35.6	50.9	990	7	W12906 ma89b03.r1
29	35.6	50.9	1006	7	W29475 mb99f11.r1
30	35.6	50.9	1020	7	W08102 mb40g08.r1
31	34.6	49.4	293	7	R47143 Y79 Rat inc
32	34.6	49.4	566	7	R46934 Y169 Rat inc
33	34.6	48.6	217	7	R46913 Y140 Rat inc
34	34.6	48.6	223	2	BB571643 BB571643
35	34.6	48.6	395	7	R46955 Y195 Rat inc
36	34.6	48.6	449	7	R47024 Y350 Rat inc
37	34.6	48.6	1031	6	CB575508 AGENCOURT
38	33.8	48.3	570	9	AY194439 Homo sapi
39	33.8	48.3	570	9	AY194439 Homo sapi
40	33.6	48.0	303	7	R46903 Y124 Rat inc
41	33.6	48.0	318	7	R47078 Y534 Rat inc
42	33.6	48.0	467	7	R46947 Y186 Rat inc
43	33.6	47.1	373	7	R47135 Y722 Rat inc
44	33.6	47.1	395	7	R47030 Y359 Rat inc
45	32.4	46.3	257	7	R47100 Y601 Rat inc

## ALIGNMENTS

RESULT 1  
CB473334/c 429 bp mRNA linear EST 26-MAR-2003  
LOCUS sn68\_E03.f sn Sus scrofa CDNA 5', mRNA sequence.  
DEFINITION CB473334  
ACCESSION CB473334  
VERSION GI:29279720  
KEYWORDS EST.  
SOURCE Sus scrofa (pig)  
ORGANISM Sus scrofa

REFERENCE 1 (bases 1 to 429)  
Neilan,J.G., Kutish,G.P., Lu,Z., Zaak,A. and Rock,D.L.  
AUTHORS Sequence analysis of African swine fever virus infected and non-infected porcine macrophage cDNA libraries  
TITLE Unpublished (2003)  
JOURNAL Contact: Neilan JG  
COMMENT Plum Island Animal Disease Center  
US Department of Agriculture, Agricultural Research Service  
PO Box 848, Greenport, NY 11944-848, USA  
Tel: 631 323 3133  
Fax: 631 323 3044  
Email: jneilane@adrc.ars.usda.gov

Single pass sequencing. Bases called with phred v0.020425.c and trimmed with the aid of the trim\_alt option. Vector identified by cross\_match v0.990329 and Lucy vl.17p.  
Seq primer: M13 Forward.  
Location/Qualifiers

1..429  
/organism="Sus scrofa"  
/mol\_type="mRNA"  
/db\_xref="taxon:9823"  
/issue\_type="lymphoid"  
/cell\_type="macrophage"  
/lab\_host="DH10B"

/clone\_lib="sn"  
/note="Vector: pSPORT1, Site 1: NotI, Site 2: SalI;  
Library made from pools of polya selected RNA, isolated at different times post-infection (0 to 16 hrs) from African swine fever virus (isolate Pretoriuskop/96/4) infected swine macrophages. Macrophages were derived from peripheral blood mononuclear cells cultured for 48 hrs on plastic in the presence of 30% U937 supernatant."

## ORIGIN

Query Match 55.1%; Score 38.6; DB 6; Length 429;  
Best Local Similarity 91.1%; Pred. No. 0.013;

Matches 41; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 GATTTGTTGCTGCTCTCGAGAGAGCCTTCAATGCTGT 45  
 |||||  
 Db 357 GATTTGTTGCTGCTCTCGAGAGAGCCTTCTATGCTCT 313

RESULT 2  
 W33932 303 bp mRNA linear EST 11-SEP-1996  
 LOCUS mb54602.r1 Soares mouse p3NMf19.5 Mus musculus cDNA clone  
 DEFINITION IMAGE:313219.5' similar to gb:M10095 Mouse amelogenin (MOUSE);

ACCESSION W33932  
 VERSION W33932.1 GI:1315837  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)

REFERENCE  
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 303)  
 Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.  
 The WashU-HMI Mouse EST Project  
 Unpublished (1996)  
 CONTACT: Marra M/Mouse EST Project  
 WashU-HMI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@wustl.edu  
 This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI:214619  
 Seq primer: mob.REGA+ET  
 High quality sequence stop: 185.  
 Location/Qualifiers

FEATURES  
 source  
 1..303  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:313219"  
 /dev\_stage="19.5 dpc total fetus"  
 /lab\_host="DH10B (ampicillin resistant)"  
 /clone\_lib="Soares mouse p3NMf19.5"  
 /note="Vector: pT73D (Pharmacia) with a modified polylinker; site\_1: Not I; site\_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTGAAGTGGAGCGCGCATTTTCTTTTCTTTT 3']  
 double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M. Fatima Bonaldo. RNA was kindly provided by Dr. Minoru Ko (Wayne State University)."

## ORIGIN

Query Match 50.9%; Score 35.6; DB 7; Length 303;  
 Best Local Similarity 90.5%; Freq. No. 0.13;  
 Matches 38; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 GATTTGTTGCTGCTCTCGAGAGAGCCTTCAATGCTGT 42  
 |||||  
 Db 81 GATTTGTTGCTGCTCTCGAGAGAGCCTTCTATGCTCT 122

RESULT 3  
 W40649

LOCUS W40649 430 bp mRNA linear EST 11-SEP-1996  
 DEFINITION mc43h12.r1 Soares mouse p3NMf19.5 Mus musculus cDNA clone  
 IMAGE:31335.5' similar to gb:M10095 Mouse amelogenin (MOUSE);  
 mRNA sequence.  
 W40649  
 ACCESSION W40649  
 VERSION W40649.1 GI:1324982  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)

REFERENCE  
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 430)  
 Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.  
 The WashU-HMI Mouse EST Project  
 Unpublished (1996)  
 CONTACT: Marra M/Mouse EST Project  
 WashU-HMI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@wustl.edu  
 This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI:223135  
 Seq primer: ETPRimer  
 High quality sequence stop: 323.  
 Location/Qualifiers

## FEATURES

source  
 1..430  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:31335"  
 /dev\_stage="19.5 dpc total fetus"  
 /lab\_host="DH10B (ampicillin resistant)"  
 /clone\_lib="Soares mouse p3NMf19.5"  
 /note="Vector: pT73D (Pharmacia) with a modified polylinker; site\_1: Not I; site\_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTGAAGTGGAGCGCGCATTTTCTTTTCTTTT 3']  
 double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M. Fatima Bonaldo. RNA was kindly provided by Dr. Minoru Ko (Wayne State University)."

## ORIGIN

Query Match 50.9%; Score 35.6; DB 7; Length 430;  
 Best Local Similarity 90.5%; Freq. No. 0.15;  
 Matches 38; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 GATTTGTTGCTGCTCTCGAGAGAGCCTTCAATGCTGT 42  
 |||||  
 Db 69 GATTTGTTGCTGCTCTCGAGAGAGCCTTCTATGCTCT 110

## RESULT 4

LOCUS A1798381 441 bp mRNA linear EST 06-JUN-1999  
 DEFINITION tr34c02.x1 NCI CGAP\_Ov23 Homo sapiens cDNA clone IMAGE:2220194.3';  
 mRNA sequence.  
 A1798381  
 ACCESSION A1798381  
 VERSION A1798381.1 GI:5363853  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 441)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
JOURNAL Tumor Gene Index  
COMMENT Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: cgaps-remail.nih.gov  
Tissue Procurement: Christopher Moshaluk, M.D., Ph.D., Michael R.  
Emmett-Buck, M.D., Ph.D.  
CDNA Library Preparation: Life Technologies, Inc.  
DNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www.bio.llnl.gov/bdip/image/image.html  
Seq primer: -40UP from Gibco  
High quality sequence stop: 395.  
Location/Qualifiers  
1..441  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2220194"  
/issue\_type="tumor, 5 pooled (see description)"  
/lab\_host="DH10B"  
/clone\_1lb="NCI-CGAP OY23"  
/note="Organ: ovary; Vector: pCMV-SPORE6; Site: 1; Salt;  
Site: 2; NotI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.35 kb. Tumor types include: mixed  
Mullerian tumor, papillary serous, clear cell, spindle  
cell. All are primary tumors, metastasis positive. Life  
Technologies catalog #: 11534-013"

ORIGIN  
Query Match 50.9%; Score 35.6; DB 1; Length 441;  
Best Local Similarity 71.2%; Pred. No. 0.15;  
Matches 47; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 4 TTGTGTCCTGCTCTCTGAGACGCTTCAGTACCTGAGTAAATTTCCATT 63  
|||||  
69 TTTCCTGACTGCTCATGAGACTAAATTCAGTTTCTGTATTAACTTACCTT 128  
|||||

QY 64 CCAATT 69  
|||

Db 129 CTCCTT 134

RESULT 5  
BE220076 444 bp mRNA linear EST 03-JUL-2000  
LOCUS hv66h03.x1 NCI-CGAP\_Lu24 Homo sapiens CDNA clone IMAGE:3178421 3',  
DEFINITION mRNA sequence.  
ACCESSION BE220076  
VERSION BE220076  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 444)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
JOURNAL Tumor Gene Index  
COMMENT Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: cgaps-remail.nih.gov  
Tissue Procurement: Christopher Moshaluk, M.D., Ph.D., Michael R.  
Emmett-Buck, M.D., Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL, send email to:  
info@image.llnl.gov  
Seq primer: -40UP from Gibco  
High quality sequence stop: 388.  
Location/Qualifiers  
1..444  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3178421"  
/issue\_type="carcinoid"  
/lab\_host="DH10B"  
/clone\_1lb="NCI-CGAP Lu24"  
/note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a  
modified polylinker; Plasmid DNA from the normalized  
library NCI-CGAP Lu5 was prepared, and as circles were  
made in vitro. Following HAP purification, this DNA was  
used as tracer in a subtractive hybridization reaction.  
The driver was PCR-amplified cDNAs from a pool of 5,000  
clones made from the same library (cloneIDB  
141920-1417991 and 1520904-1522439). Subtraction by Bento  
Soares and M. Fatima Bonaldo."

ORIGIN  
Query Match 50.9%; Score 35.6; DB 2; Length 444;  
Best Local Similarity 71.2%; Pred. No. 0.15;  
Matches 47; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 4 TTGTGTCCTGCTCTCTGAGACGCTTCAGTACCTGAGTAAATTTCCATT 63  
|||||  
65 TTTCCTGACTGCTCATGAGACTAAATTCAGTTTCTGTATTAACTTACCTT 124  
|||||

QY 64 CCAATT 69  
|||

Db 125 CTCCTT 130

RESULT 6  
A1338491 448 bp mRNA linear EST 13-FEB-1999  
LOCUS q997h05.x1 Soares total fetus Nb2HF8\_9w Homo sapiens CDNA clone  
DEFINITION IMAGE:193953 3', mRNA sequence.  
ACCESSION A1338491  
VERSION A1338491  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 448)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
JOURNAL Tumor Gene Index  
COMMENT Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: cgaps-remail.nih.gov  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert length: 1345 Std Error: 0.00  
Seq primer: -40UP from Gibco  
High quality sequence stop: 447.  
Location/Qualifiers  
1..448  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:193953"  
/dev\_stage="8-9 weeks"  
/lab\_host="DH10B"  
/clone\_1lb="Soares total fetus Nb2HF8\_9w"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker; Site: 1; Not I; Site: 2; Eco RI; 1st strand cDNA  
was prepared from mRNA obtained from pooled 8-9 week

## ORIGIN

(total) fetus material with a Not I - oligo (dT) primer [5' TGTTACCATCTAGTGGAGGCGCGCTTAATTTTCTTTT 3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo. "

Query Match 50.9%; Score 35.6; DB 1; Length 448;  
Best Local Similarity 71.2%; Pred. No. 0.15;  
Matches 47; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 4 TTGTTGCTGCTGCTGCTGAGAGCCTTACAGTACCCAGTAAATTTCCAAATT 63  
|||||  
DB 69 TTTCCTGACTGCTCATGAGAGCTAAATTCAGTTTCCGTATTAACATTACCTTT 128  
|||||  
QY 64 CCAATT 69  
|||  
DB 129 CTCCTT 134

RESULT 7  
W36345 536 bp mRNA linear EST 11-SEP-1996  
LOCUS M272C12.r1 Soares mouse p3NM19.5 Mus musculus cDNA clone  
DEFINITION IMAGE:334966 5' similar to gp:M10095 Mouse amelogenin (MOUSE) ;  
mRNA sequence.

ACCESSION W36345.1 GI:1318120  
VERSION W36345  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 536)  
AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Gaisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, B., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Tietelng, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.

TITLE The WashU-HMI Mouse EST Project  
JOURNAL Unpublished (1996)  
COMMENT Contact: Marra M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.edu  
This clone is available royalty-free through LBNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:216366

FEATURES  
source High quality sequence stop: 359.  
Location/Qualifiers  
1..536  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:334966"  
/dev\_stage="19.5 dpc total fetus"  
/lab\_host="DH10B (ampicillin resistant)"  
/clone\_id="Soares mouse p3NM19.5"  
/note="Vector: pT73D (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCATCTAGTGGAGGCGCGCTTAATTTTCTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento

## ORIGIN

Soares and M. Fatima Bonaldo. RNA was kindly provided by Dr. Minoru Ko (Wayne State University). "

Query Match 50.9%; Score 35.6; DB 7; Length 536;  
Best Local Similarity 90.5%; Pred. No. 0.15;  
Matches 38; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GATTTTGTGCTGCTGCTGCTGAGAGCCTTACAGTATGCC 42  
|||||  
DB 64 GATTTTGTGCTGCTGCTGCTGAGAGCCTTTTGTCTATGCC 105  
|||||

RESULT 8  
BX470683/c 561 bp mRNA linear EST 04-SEP-2003  
LOCUS DKFZP686N21110.r1 686 (synonym: h1cc3) Homo sapiens cDNA clone  
DEFINITION DKFZP686N21110 5', mRNA sequence.  
ACCESSION BX470683  
VERSION BX470683.1 GI:31665009  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 561)  
AUTHORS Koehrer, K., Beyer, A., Mewes, H.W., Well, B., Amid, C., Oeanger, A., Fodor, G., Han, W., and Wiemann, S.  
EST (Koehrer, K., Beyer, A., Mewes, H.W., Well, B., Amid, C., et al.)  
Unpublished (2003)  
Contact: MIPS

MIPS Ingolstaedter Landstr.1, D-85764 Neuberg, Germany  
This is the 5' sequence of the clone insert.  
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;  
sequenced by BMFZ (Biomedical Research Center at the Heinrich-Heine-University, Dusseldorf/Germany) within the cDNA sequencing consortium of the German Genome Project. No 81 sequence available.  
This clone (DKFZP686N21110) is available at the RZPD in Berlin.  
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

FEATURES  
source 1..561  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="DKFZP686N21110"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/clone\_id="686 (synonym: h1cc3)"  
/note="Vector: pT73D; Site 1: SfiI; Site 2: SfiI; cDNA-collection"

## ORIGIN

Query Match 50.9%; Score 35.6; DB 5; Length 561;  
Best Local Similarity 71.2%; Pred. No. 0.15;  
Matches 47; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 4 TTGTTGCTGCTGCTGCTGAGAGCCTTACAGTATGCTGTGAGTAAATTTCCAAATT 63  
|||||  
DB 211 TTTCCTGACTGCTCATGAGAGCTAAATTCAGTTTCCGTATTAACATTACCTTT 152  
|||||  
QY 64 CCAATT 69  
|||  
DB 151 CTCCTT 146

RESULT 9  
AY419441 585 bp DNA linear GSS 17-DEC-2003  
LOCUS AY419441  
DEFINITION Mus musculus AMELX gene, VIRTUAL TRANSCRIPT, partial sequence, genomic survey sequence.



ACCESSION AY419441  
 VERSION AY419441.1 GI:39775398  
 KEYWORDS GSS.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 REFERENCE 1 (bases 1 to 585)  
 AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarimal,A., Todd,M.A., Tanenbaum,D.M., Civallo,D.R., Lu,F., Murphy,B., Ferriere,S., Wang,G., Zheng,X.H., White,T.J., Sinsky,J.J., Adams,M.D. and Cargill,M.  
 TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios  
 JOURNAL Science 302 (5652), 1960-1963 (2003)  
 PUBMED 14671302  
 REFERENCE 2 (bases 1 to 585)  
 AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarimal,A., Todd,M.A., Tanenbaum,D.M., Civallo,D.R., Lu,F., Murphy,B., Ferriere,S., Wang,G., Zheng,X.H., White,T.J., Sinsky,J.J., Adams,M.D. and Cargill,M.  
 TITLE Direct Submision  
 JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA  
 COMMENT These sequences were made by sequencing genomic exons and ordering them based on alignment.  
 FEATURES  
 source Location/Qualifiers  
 1..585  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10090"  
 <1..>585  
 /gene="AMELX"  
 /locus\_tag="HGM6891"

ORIGIN  
 Query Match 50.9%; Score 35.6; DB 9; Length 585;  
 Best Local Similarity 90.5%; Pred. No. 0.16;  
 Matches 38; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GATTTTGTTCCTGCTGCTGAGAGAGCCTTCAGTATGCC 42  
 12 GATTTTGTTCCTGCTGCTGAGAGAGCCTTCAGTATGCC 53

Db 12 GATTTTGTTCCTGCTGCTGAGAGAGCCTTCAGTATGCC 53

RESULT 10  
 BF001849 586 bp mRNA linear EST 06-OCT-2000  
 LOCUS 7994c09.x1 NCI\_CGAP\_Col6 Homo sapiens cDNA clone IMAGE:3314128 3',  
 DEFINITION mRNA sequence.  
 ACCESSION BF001849  
 VERSION BF001849.1 GI:10702124  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 REFERENCE 1 (bases 1 to 586)  
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-remail.nih.gov  
 Tissue Procurement: Ilian Kirsch, M.D., Michael R. Emmert-Buck, M.D., Ph.D.  
 CDNA Library Preparation: M. Bento Soares, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL, send email to: info@image.llnl.gov  
 Seg primer: -40UP from Gibco

FEATURES  
 source High quality sequence stop: 471.  
 Location/Qualifiers  
 1..586  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3314128"  
 /tissue\_type="colon tumor. RER+"  
 /lab\_host="DH10B"  
 /clone\_1lb="NCI CGAP Col6"  
 /note="Organ: colon; Vector: pTT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not 1; Site 2: Eco RI; Plasmid DNA from the normalized library NCI CGAP Col6 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneids 1057416-1061255, and 114584-1145351).  
 Subtraction by Bento Soares and M. Fatima Bonaldo. "

ORIGIN  
 Query Match 50.9%; Score 35.6; DB 2; Length 586;  
 Best Local Similarity 71.2%; Pred. No. 0.16;  
 Matches 47; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 4 TTTGTTGCTGCTGCTGAGAGACCTTCAGTATGCTGAGTAAATTCGATTT 63  
 65 TTTCCTGACTGCTCATGAGAGCTAAATTCAGTTTCTGTATTAACATTTACCTT 124

Db 64 CCAATT 69  
 125 CTCCTT 130

RESULT 11  
 BF448291 588 bp mRNA linear EST 01-DEC-2000  
 LOCUS nad17f04.x1 NCI\_CGAP\_Lu24 Homo sapiens cDNA clone IMAGE:336587 3',  
 DEFINITION mRNA sequence.  
 ACCESSION BF448291  
 VERSION BF448291.1 GI:11513814  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 REFERENCE 1 (bases 1 to 588)  
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-remail.nih.gov  
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
 CDNA Library Preparation: M. Bento Soares, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL, send email to: info@image.llnl.gov  
 Seg primer: -40UP from Gibco  
 High quality sequence stop: 465.  
 Location/Qualifiers  
 1..588  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:336587"  
 /tissue\_type="carcinoid"  
 /lab\_host="DH10B"  
 /clone\_1lb="NCI CGAP Lu24"  
 /note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a

## ORIGIN

modified polylinker; plasmid DNA from the normalized library NCI CGAP-Lus was prepared, and as circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneids 1414920-1417991 and 1520904-1522439). Subtraction by Bento Soares and M. Fatima Bernaldo.

Query Match 50.9%; Score 35.6; DB 2; Length 588;

Best Local Similarity 71.2%; Pred. No. 0.16; Mismatches 19; Indels 0; Gaps 0;

Matches 47; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 4 TTTGTTGCTGCTCTCTGGAGACCTTCAGTATGCTGTGAGTAAATTTCCAAATT 63  
 DB 69 TTTCTTGAAGCTGCTCATGAGACTTAATTCAGTTTCTGTATTAACATTACACTTT 128  
 QY 64 CCAATT 69  
 DB 129 CTCCTT 134

RESULT 12  
 BG622884/c 620 bp mRNA linear EST 18-APR-2001  
 LOCUS 602647718F1 NIH\_MGC\_79 Homo sapiens cDNA clone IMAGE:4769213 5',  
 DEFINITION mRNA sequence.  
 ACCESSION BG622884  
 VERSION BG622884.1 GI:13674255  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 620)  
 NIH-MGC http://mgs.nci.nih.gov/.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 CONTACT: Robert Strauberg, Ph.D.  
 Email: cgaabs-remail.nih.gov  
 Tissue Procurement: CLONTECH Laboratories, Inc.  
 CDNA Library Preparation: CLONTECH Laboratories, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Incey Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.jnl.gov  
 Plate: LNCM634 row: c column: 06  
 High quality sequence stop: 618.

## FEATURES

## Source

1. 620  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /mol\_type="RNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4769213"  
 /lab\_host="DH10B (TI phage-resistant)"  
 /clone\_id="NIH\_MGC\_79"  
 /note="Organ: placenta; Vector: pDNR-LIB (Clontech); Site 1: SfiI (ggcgatcggc); Site 2: SfiI (ggcgatcggc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGGCCATTATGCGC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGGCGCGCAGCATG-dt(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.3 kb (range 0.5-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 50.9%; Score 35.6; DB 4; Length 620;  
 Best Local Similarity 71.2%; Pred. No. 0.16;

Matches 47; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 4 TTTGTTGCTGCTCTCTGGAGACCTTCAGTATGCTGTGAGTAAATTTCCAAATT 63  
 DB 513 TTTCTTGAAGCTGCTCATGAGACTTAATTCAGTTTCTGTATTAACATTACACTTT 454

QY 64 CCAATT 69  
 DB 453 CTCCTT 448

## RESULT 13

CD773419 645 bp mRNA linear EST 02-JUL-2003

LOCUS CD773419 IMAGE:30501267 5', mRNA sequence.

DEFINITION CD773419.1 GI:32431921

ACCESSION CD773419

VERSION CD773419.1 GI:32431921

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 645)  
 NIH-MGC http://mgs.nci.nih.gov/.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 CONTACT: Daniela S. Gerhard, Ph.D.  
 Office of Cancer Genomics  
 National Cancer Institute / NIH  
 Bldg. 31 Rm10A07 Bethesda, MD 20892  
 Email: cgaabs-remail.nih.gov  
 Tissue Procurement: Yoshiniko Yamada, Takashi Nakamura, NIDCR  
 CDNA Library Preparation: CLONTECH Laboratories, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.jnl.gov  
 Plate: NDCM223 row: k column: 04  
 High quality sequence stop: 611.

## FEATURES

## Source

1. 645  
 Location/Qualifiers  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:30501267"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_id="NIH\_MGC\_190"  
 /note="Organ: Pooled - Molax; Vector: pDNR-LIB; Site 1: SfiI (ggcgatcggc); Site 2: SfiI (ggcgatcggc); Non-normalized full-length enriched library 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGGCCATTATGCGC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGGCGCGCAGCATG-dt(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.71 kb (range 0.5-3.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA) Corp."

## ORIGIN

Query Match 50.9%; Score 35.6; DB 6; Length 645;

Best Local Similarity 90.5%; Pred. No. 0.16; Mismatches 4; Indels 0; Gaps 0;

Matches 38; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GATTTGTTGCTGCTCTCTGGAGACCTTCAGTATGCGC 42  
 DB 84 GATTTGTTGCTGCTCTCTGGAGACCTTCAGTATGCGC 125

RESULT 14  
 CB056709

LOCUS CB056709 707 bp mRNA linear EST 17-JAN-2003  
 DEFINITION NISC J119905.w1 Soares NMBP13-15 Mus musculus cDNA clone  
 IMAGE:4848584 5', mRNA sequence.  
 ACCESSION CB056709  
 VERSION CB056709.1 GI:27794996  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 707)  
 REFERENCE NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 TITLES Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT Contact: Robert Straubeberg, Ph.D.  
 Email: [CGAP@rs-romail.nih.gov](mailto:CGAP@rs-romail.nih.gov)  
 CDNA Library Preparation:  
 DNA Sequencing by: The I.M.A.G.E. Consortium/LNL  
 Sequencing Center (NISC)  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
[info@image.llnl.gov](mailto:info@image.llnl.gov)  
 Plate: L1AM10779 row: N column: 9  
 Seq primer: T7 primer  
 Location/Qualifiers  
 1..707  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:4848584"  
 /tissue\_type="pituitary gland"  
 /dev\_stage="juvenile, 13-15 days"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_1lb="Soares NMBP13-15"  
 /note="Organ: brain; Vector: pTZ19-PacI; Site 1: NotI;  
 Site 2: EcoRI; 1st strand cDNA was primed with a NotI-  
 oligo(dT) primer  
 5'-ACCTGAGAGATTGCGCGCGCTGACCGATGTTTCTTTTCTTTT-3';  
 double-stranded cDNA was ligated to EcoRI adaptors  
 5'-ATTGCGACGACGAG-3' and 5'-CTCGTCCG-3' (Pharmacia),  
 digested with NotI and cloned into the NotI and EcoRI  
 sites of the pTZ19-PacI vector. Library went through one  
 round of normalization, and was constructed in the  
 laboratory of M. Bento Soares (University of Iowa)."

ORIGIN

Query Match 50.9%; Score 35.6; DB 6; Length 707;  
 Best Local Similarity 90.5%; Pred. No. 0.16;  
 Matches 38; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GATTTGTTTGCTGCTGCTGAGAGACGCTTCAGTATGCC 42  
 |||||  
 DB 58 GATTTGTTTGCTGCTGCTGAGAGACGCTTTTGTATGCC 99  
 |||||

RESULT 15  
 CN360063 707 bp mRNA linear EST 16-MAY-2004  
 LOCUS 17000583090269 GRN\_PREHEP Homo sapiens cDNA 5', mRNA sequence.  
 DEFINITION CN360063  
 ACCESSION CN360063  
 VERSION CN360063.1 GI:47359997  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 707)  
 REFERENCE Brandenberger, R., Wei, H., Zhang, S., Lei, S., Murae, J., Fisk, G.J.,  
 Li, Y., Xu, C., Fang, R., Guegler, K., Rao, M.S., Mandalam, R.,  
 Lebkowski, J. and Stanton, L.W.  
 Transcriptional characterization elucidates signaling networks that

JOURNAL control human ES cell growth and differentiation  
 Nat. Biotechnol. 22 (6), 707-716 (2004)  
 COMMENT Contact: Brandenberger R  
 Regenerative Medicine  
 Genon Corporation  
 230 Constitution Drive, Menlo Park, CA 94025, USA  
 Tel: 650 473 8658  
 Fax: 650 473 7760  
 Email: [rbrandenberger@genon.com](mailto:rbrandenberger@genon.com)  
 Insert Length: 707 Std Error: 0.00.  
 Location/Qualifiers  
 1..707  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /tissue\_type="embryonic stem cells, DMSO-treated H9 cell  
 line"  
 /clone\_1lb="GRN PREHEP"  
 /note="oligo dT primed, full-length enriched cDNA library  
 from DMSO-treated hES cell line H9 (p22) maintained in  
 feeder-free conditions".

ORIGIN

Query Match 50.9%; Score 35.6; DB 7; Length 707;  
 Best Local Similarity 71.2%; Pred. No. 0.16;  
 Matches 47; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 4 TTTGTTGCTGCTGCTGAGAGACGCTTCAGTATGCTGTAGTAAATTTCCAATTT 63  
 |||||  
 DB 596 TTTGTTGCTGCTGCTGAGAGACGCTTCAGTATGCTGTATTAACATTTACACTTT 537  
 |||||

QY 64 CCAATT 69  
 |||||  
 DB 536 CTCCTT 531  
 |||||

Search completed: February 8, 2005, 22:18:08  
 Job time : 835.766 sec

**This Page Blank (uspto)**



## ORIGIN

/translation="MGTWILFASLGAAPAIPLPSPHPIGNTSYEVLTPLKRYOS  
LIRQPTSYGYEPMGMLHRIIIVLSQNPNSHALQPHHILPMVSAQHPVVPQCPM  
PLPGSHVPTQHNQPNLPPVQOPPHPOVOPPOPPPOPLPPIOPPLPPLPPLP  
PIFPIQPLPPLPDLPLEAMPATG"

Query Match 71.3%; Score 47.8; DB 4; Length 5712;  
Best Local Similarity 82.1%; Pred. No. 8e-07;  
Matches 55; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 1 CACAAGCTTGAATAAAGTCTGCCACCAGTTGTAAGGTTTAAGACAGTA 60  
|||||  
Db 1669 CACAAGCTTGAATAAATCTCTCCACGGTTGTAACCTTAAGGTTAAGACAGTA 1728  
|||||

Qy 61 CAAGATC 67  
|||||  
Db 1729 CAAGATC 1735

RESULT 2  
AB091789 6451 bp DNA linear MAM 02-MAY-2003  
LOCUS  
DEFINITION Bos taurus AMELX gene for amelogenin, partial cds.  
ACCESSION AB091789  
VERSION AB091789.1 GI:29126030  
KEYWORDS  
SOURCE Bos taurus (cow)  
ORGANISM Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovinae; Bos.

REFERENCE  
AUTHORS Iwase, M., Satta, Y., Hirai, Y., Hirai, H., Imai, H. and Takahata, N.  
TITLE From the Cover: The amelogenin loci span an ancient pseudautosomal  
boundary in diverse mammalian species  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 100 (9), 5258-5263 (2003)  
MEDLINE 12672962  
PUBMED 12672962

REFERENCE  
AUTHORS Iwase, M., Satta, Y. and Takahata, N.  
TITLE Direct Submission  
JOURNAL Submitted (19-SEP-2002) Mineyo Iwase, Graduate University for  
Advanced Studies (Sokendai), Department of Biosystems Science;  
Shonan Kokusaiinura, Hayama, Kanagawa 240-0193, Japan  
(E-mail: iwase@min.koryuwo1.soken.ac.jp, Tel: 81-468-58-1571,  
Fax: 81-468-58-1544)  
FEATURES  
source Location/Qualifiers  
1..6451  
/organism="Bos taurus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9913"  
/sex="male"  
238..6451  
/gene="AMELX"  
join(238..340,1650..1661)  
/gene="AMELX"  
join(1662..1715,4173..4220,5560..5601,5665..5709,  
5987..56451)  
/gene="AMELX"  
/codon\_start=1  
/product="amelogenin"  
/protein\_id="BAC66108.1"  
/db\_xref="GI:29126031"

Qy 5'UTR  
CDS  
ORIGIN  
Query Match 71.3%; Score 47.8; DB 4; Length 6451;  
Best Local Similarity 82.1%; Pred. No. 8e-07;  
Matches 55; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 1 CACAAGCTTGAATAAAGTCTGCCACCAGTTGTAAGGTTTAAGACAGTA 60  
|||||  
Db 1755 CACAAGCTTGAATAAAGTCTGCCACCAGTTGTAAGGTTTAAGACAGTA 1814  
|||||

Qy 61 CAAGATC 67  
|||||  
Db 1815 CATGATC 1821

RESULT 3  
AB091785 5684 bp DNA linear PRI 02-MAY-2003  
LOCUS  
DEFINITION Lemur catta AMELX gene for amelogenin, partial cds.  
ACCESSION AB091785  
VERSION AB091785.1 GI:29126023  
KEYWORDS  
SOURCE Lemur catta (ring-tailed lemur)  
ORGANISM Lemur catta  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Strepsithini; Lemniridae; Lemur.

REFERENCE  
AUTHORS Iwase, M., Satta, Y., Hirai, Y., Hirai, H., Imai, H. and Takahata, N.  
TITLE From the Cover: The amelogenin loci span an ancient pseudautosomal  
boundary in diverse mammalian species  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 100 (9), 5258-5263 (2003)  
MEDLINE 12672962  
PUBMED 12672962

REFERENCE  
AUTHORS Iwase, M., Satta, Y. and Takahata, N.  
TITLE Direct Submission  
JOURNAL Submitted (19-SEP-2002) Mineyo Iwase, Graduate University for  
Advanced Studies (Sokendai), Department of Biosystems Science;  
Shonan Kokusaiinura, Hayama, Kanagawa 240-0193, Japan  
(E-mail: iwase@min.koryuwo1.soken.ac.jp, Tel: 81-468-58-1571,  
Fax: 81-468-58-1544)  
FEATURES  
source Location/Qualifiers  
1..5684  
/organism="Lemur catta"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9447"  
/sex="male"  
100..5684  
/gene="AMELX"  
join(100..156,1503..1514)  
/gene="AMELX"  
join(1515..1568,3465..3512,4804..4845,4937..4981,  
5229..5684)  
/gene="AMELX"  
/codon\_start=1  
/product="amelogenin"  
/protein\_id="BAC66105.1"  
/db\_xref="GI:29126024"

Qy 1 CACAAGCTTGAATAAAGTCTGCCACCAGTTGTAAGGTTTAAGACAGTA 60  
|||||  
Db 1608 CACAAGCTTGAATAAATCTCTCCACGGTTGTAACCTTAAGGTTAAGACAGTA 1667  
|||||

Qy 61 CAAGATC 67  
|||||  
Db 1668 TAAATTC 1674

RESULT 4  
Query Match 66.6%; Score 44.6; DB 9; Length 5684;  
Best Local Similarity 79.1%; Pred. No. 1.2e-05;  
Matches 53; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

LOCUS	AB091787	5562 bp	DNA	linear	PRI 02-MAY-2003
DEFINITION	Otolemur garnettii AMELX gene for amelogenin, partial cds.				
ACCESSION	AB091787				
VERSION	AB091787.1	GI:29126027			
KEYWORDS					
SOURCE	Otolemur garnettii (small-eared galago)				
ORGANISM	Otolemur garnettii				
REFERENCE					
AUTHORS	1				
TITLE	Iwase, M., Satta, Y., Hirai, Y., Hirai, H., Imai, H., and Takahata, N.				
JOURNAL	From the Cover: The amelogenin loci span an ancient pseudoautosomal boundary in diverse mammalian species				
PUBLISHED	Proc. Natl. Acad. Sci. U.S.A. 100 (9), 5258-5263 (2003)				
REFERENCE	2 (bases 1 to 5562)				
TITLE	Iwase, M., Satta, Y. and Takahata, N.				
JOURNAL	Direct Submission				
AUTHORS	Submitted (19-Sep-2002) Mineyo Iwase, Graduate University for Advanced Studies (Sokendai), Department of Biosystems Science, Shonan Kikusaijima, Hayama, Kanagawa 240-0193, Japan				
REFERENCE	(E-mail: iwase@shonan.kyushu-u.ac.jp, Tel: 81-468-58-1571, Fax: 81-468-58-1544)				
FEATURES					
source	Location/Qualifiers				
	1..5562				
	/organism="Otolemur garnettii"				
	/mol_type="genomic DNA"				
	/db_xref="taxon:30611"				
	/sex="male"				
gene	267..5562				
	/gene="AMELX"				
5'UTR	join(267..321,1401..1412)				
	/gene="AMELX"				
CDS	join(1413..1466,3385..3432,4699..4740,4832..4876,5125..5562)				
	/gene="AMELX"				
	/codon_start=1				
	/product="amelogenin"				
	/protein_id="BAC66107.1"				
	/db_xref="GI:29126028"				
	/translation="MGTVITLPACLGAAPMLPPHGHGPEYINFSYEKYSQAINIDRTALVTVKXVQSNLRPRPSYSGEPPGMLHHQILPVLSQNPRTILQPHNRIIVVPADQPVVPOQPMVVPQGHSMPTLQNHQPMVLPRAQDPQVQDPQDPQMPQMPQMPQMPQIPQDPQPVHPMQPLPQPPPLPPLPMPQPLPMLDLEAMPAT"				
ORIGIN					
	Query Match 59.4% Score 39.8; DB 9; Length 5562;				
	Best Local Similarity 74.6% Pred. No. 0.0075;				
	Matches 50; Conservative 0; Mismatches 17; Indels 0; Gaps 0;				
OY	1 CACAAGCTTGAATAAAGTCTGCCACACAGTGTGTAAGTTAAGGTTAAGACAGTA 60				
DB	1506 CATTAAGCTTAATAAATAAATCTGCTCAACAGTGTGTAACCTCAGATTAAACAGTA 1565				
OY	61 CAAGATC 67				
DB	1566 TAGTATC 1572				
RESULT 5					
LOCUS	AB091792	6931 bp	DNA	linear	MAV 02-MAY-2003
ACCESSION	AB091792				
VERSION	AB091792.1	GI:29126036			
KEYWORDS					
SOURCE	Sus scrofa (pig)				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.				

FEATURES	source
gene	5'UTR
CDS	
ORIGIN	
Query Match	55.5%; Score 37.2; DB 4; Length 6931;
Best Local Similarity	72.7%; Pred. No. 0.0069;
Matches	48; Conservative 0; Mismatches 18; Indels 0; Gaps 0;
Oy	1 CACAACTTGGAAATAAAGTCGCCGCCACGAGTGTAACTTTAGGCTTTAAGACAGTA 60
Db	2187 CATTAAGCTTGGAAATAAATAATTCGCCACATAGATGTAACTTCAGGCTTTAAACAGAA 2246
Oy	61 CAAGAT 66
Db	2247 CCAGAT 2252
RESULT 6	
AB091781	6465 bp DNA linear PRI 02-MAY-2003
LOCUS	AB091781
DEFINITION	Pan troglodytes AMELX gene for amelogenin, partial cds.
ACCESSION	AB091781
VERSION	AB091781.1 GI:29126015
KEYWORDS	
SOURCE	Pan troglodytes (chimpanzee)
ORGANISM	Pan troglodytes
REFERENCE	Eukaryote; Metacoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.
AUTHORS	Iwase,M., Satta,Y., Hirai,Y., Hirai,H., Inai,H. and Takahata,N.
TITLE	From the Cover: The amelogenin loci span an ancient pseudautosomal boundary in diverse mammalian species
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 100 (9), 5258-5263 (2003)
MEDLINE	22608569
PUBMED	12672962
REFERENCE	2 (bases 1 to 6931)
AUTHORS	Iwase,M., Satta,Y. and Takahata,N.
TITLE	Direct Submission
JOURNAL	Submitted (19-SEP-2002) Mineyo Iwase, Graduate University for Advanced Studies(Sokendai), Department of Biosystems Science; Shonan kokuuimura, Hayama, Kangawa 240-0193, Japan
REFERENCE	(E-mail:iwase@koryu01.soken.ac.jp, Tel:81-468-58-1571, Fax:81-468-58-1544)
FEATURES	Location/Qualifiers
source	1..6931
organism	"Sus scrofa"
mol_type	"genomic DNA"
db_xref	"taxon:9823"
sex	"male"
gene	"AMELY"
join	(625..680,2096..2107)
gene	"AMELY"
join	(2108..2161,4806..4853,6120..6161,6218..6262,6534..>6931)
gene	"AMELY"
product	"amelogenin"
protein_id	"BAC6611.1"
db_xref	"GI:29126037"
translation	"MGTWIFACLLGASLAMP,PPHPGAGYINFSYEDLYLEAICIDRTAFVLPLKMYONMIRPYTSGYSEBFGMLHQLIPVSOQPSHALQPHNIPMVPAOQGPRIPOPMPLPGQSMTPTOHQNPLPLPAQGPQGPVQPHQPLQGPSPMHPIQRLPQPLPMPMSQSLPLDLLEWMPATD"

Shonan kokusaiimura, Hayama, Kanagawa 240-0193, Japan  
[E-mail: iwaseem@koryu.w01.soken.ac.jp, Tel: 81-468-58-1571,  
Fax: 81-468-58-1544]

## FEATURES

source

Location/Qualifiers

1..6465  
/organism="Pan troglodytes"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9598"  
/sex="male"  
927..6465  
/gene="AMELX"  
join(927..982,2283..2294)  
/gene="AMELX"  
join(2295..2348,4284..4331,5607..5648,5740..5781,  
6052..56465)  
/gene="AMELX"  
/codon\_start=1  
/product="amelogenin"  
/protein\_id="BAC66101.1"  
/db\_xref="GI:29126016"  
/translation="MGTWILFACLLGAAPFNPVPHGPGYINSEYSHQATIND  
RTALVLTPLKWTQIRPYSTGYEPMGWLHQTIPVSOQPHPTLQPHNHLIV  
PAAQGVIPQOPMPLPGOSHMTPIQHNQPNLPLPAQOPFOPOPOPOPOPPV  
HPWQPLPQPLPMPFPMQPLPPLPDLTLEAMPSTDKTK"

## ORIGIN

Query Match 55.2%; Score 37; DB 9; Length 6465;  
Best Local Similarity 82.1%; Pred. No. 0.0082;  
Matches 55; Conservative 0; Mismatches 10; Indels 2; Gaps 1;

QY 1 CACAGCTTGAATAAAGTCTGCCACCACTTGTAGGTTTAAAGACAGTA 60  
DB 2388 CACAGCTTGAATAAATCTGCCCC-ATACTGTGTAATTAAGCTTTAAACAGTA 2445  
QY 61 CAAGATC 67  
DB 2446 TGAGATC 2452

RESULT 7  
AB091791 7425 bp DNA linear MAM 02-MAY-2003  
LOCUS  
DEFINITION  
AB091791  
AB091791.1 GI:29126034  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Sus scrofa (pig)  
Sus scrofa  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
Iwase, M., Satta, Y., Hirai, Y., Hirai, H., Imai, H. and Takahata, N.  
From the Cover: The amelogenin loci span an ancient pseudautosomal  
boundary in diverse mammalian species  
Proc. Natl. Acad. Sci. U.S.A. 100 (9), 5258-5263 (2003)  
12672962  
2 (bases 1 to 7425)  
Iwase, M., Satta, Y. and Takahata, N.  
Direct Submission  
Submitted (19-SEP-2002) Mineyo Iwase, Graduate University for  
Advanced Studies (Soken), Department of Biosystems Science;  
Shonan kokusaiimura, Hayama, Kanagawa 240-0193, Japan  
[E-mail: iwaseem@koryu.w01.soken.ac.jp, Tel: 81-468-58-1571,  
Fax: 81-468-58-1544]

## FEATURES

source

Location/Qualifiers

1..7425  
/organism="Sus scrofa"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9823"  
/sex="male"  
1491..7425  
/gene="AMELX"

5'UTR join(1491..1548,3140..3151)  
/gene="AMELX"  
join(3152..3205,5316..5363,6619..6660,6717..6761,  
7033..57425)  
CDS  
/gene="AMELX"  
/codon\_start=1  
/product="amelogenin"  
/protein\_id="BAC66110.1"  
/db\_xref="GI:29126035"  
/translation="MGTWILFACLLGAAPFNPVPHGPGYINSEYEDLYLEAIRD  
RTALVLTPLKWTQIRPYSTGYEPMGWLHQTIPVSOQPHPTLQPHNHLIV  
PAAQGVIPQOPMPLPGOSHMTPIQHNQPNLPLPAQOPFOPOPOPOPOPPV  
HPWQPLPQPLPMPFPMQPLPPLPDLTLEAMPSTDKTK"

## ORIGIN

Query Match 54.6%; Score 36.6; DB 4; Length 7425;  
Best Local Similarity 71.6%; Pred. No. 0.012;  
Matches 48; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 1 CACAGCTTGAATAAAGTCTGCCACCACTTGTAGGTTTAAAGACAGTA 60  
DB 3245 CATTAGCTTGAATAAATCTACCCGAGTTGATACAGTTAGGATGAACAGTA 3304  
QY 61 CAAGATC 67  
DB 3305 GAAGATC 3311

RESULT 8  
DB3729 2501 bp DNA linear PRI 10-JUL-2001  
LOCUS  
DEFINITION  
DB3729  
DB3729.1 GI:5263178  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
Sekiuchi, H., Minaguchi, K., Machida, Y. and Yakushiji, M.  
PCR detection of the human amelogenin gene and its application to  
the diagnosis of amelogensis imperfecta  
Bull. Tokyo Dent. Coll. 39 (4), 275-285 (1998)  
99234629  
10218009  
2 (bases 1 to 2501)  
Sekiuchi, H.  
Direct Submission  
Submitted (29-FEB-1996) Hiroshi Sekiuchi, Tokyo Dental College,  
Pediatric Dentistry; 1-2-2 Maeno, Minami-ku, Chiba 261,  
Japan [Tel: 043-270-3945, Fax: 043-279-2052]

## FEATURES

source

Location/Qualifiers

1..2501  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="Y"  
1..2501  
/gene="AMGY"  
1..2501  
/note="AMGY"  
this nucleotide sequence shows an intron between exon2 and  
exon3 within human amelogenin gene on the Y chromosome"  
/number=2

## ORIGIN

Query Match 54.0%; Score 36.2; DB 9; Length 2501;  
Best Local Similarity 72.3%; Pred. No. 0.016;  
Matches 47; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 2 ACAAGCTTGAATAAAGTCTGCCACCACTTGTAGGTTTAAAGACAGTAC 61



Db 39 ACBAAGCTGCAATTAAGATTGTCCTCCGACGCGCAACTTAAGCTTTAAACAGCTT 98  
QY 62 AGAT 66  
Db 99 GAGCT 103

RESULT 9  
AP000918/c 177654 bp DNA linear HTG 30-MAY-2000  
LOCUS Homo sapiens chromosome 18 clone RP11-737C10 map 18p11.3, WORKING  
DEFINITION DRAFT SEQUENCE, 27 unordered pieces.  
ACCESSION AP000918.3 GI:8119059  
VERSION HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 177654) Toyoda, A., Taylor, T. D., Hong-Seog, P.,  
Hattori, M., Ishii, K., Yoda, T., Totoki, Y., Watanabe, H. and Sakaki, Y.  
Fujiyama, A., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.  
Homo sapiens 177,654 genomic DNA of 18p11.3  
Published Only in Database (1999)  
2 (bases 1 to 177654) Hattori, M., Ishii, K., Toyoda, A., Taylor, T. D., Hong-Seog, P.,  
Fujiyama, A., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.  
Direct Submission  
Submitted (17-DEC-1999) Masahira Hattori, The Institute of Physical  
and Chemical Research (RIKEN), Genomic Sciences Center (GSC),  
Katsuo Univ., 1-15-1 Katsuo, Sagamihara, Kanagawa 228-8555,  
Japan (E-mail:hattori@gs.c.riken.go.jp,  
URL:http://hgp.gs.c.riken.go.jp/, Tel:81-42-778-9923,  
Fax:81-42-778-9924)  
On May 31, 2000 this sequence version replaced gi:6997734.

----- Genome Center  
Center: RIKEN Genomic Sciences Center (GSC)  
Center code: RIKEN  
Web site: http://hgp.gs.c.riken.go.jp/  
Contact: hattori@gs.c.riken.go.jp  
----- Project Information  
Center project name: HumDraet2  
Center clone name: Summary Statistics  
Sequencing vector: PCR products; 100% of reads  
Chemistry: Dye-terminator ET-amersham; 100% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 15691 bases at least Q40  
Consensus quality: 157448 bases at least Q30  
Consensus quality: 17231 bases at least Q20  
Insert size: 175054; sum-of-coverage  
Quality coverage: 4.38x in Q20 bases; sum-of-coverage

-----  
NOTE: This is a 'working draft' sequence. It currently consists of  
27 contigs. The true order of the pieces is not known and their  
order in this sequence record is arbitrary. Gaps between the  
contigs are represented as runs N, but the exact sizes of the gaps  
are unknown. This record will be updated with the finished sequence  
as soon as it is available and the accession number will be  
preserved

1 46938 contig of 46938 bp in length  
47039 69628 contig of 22590 bp in length  
69729 77803 contig of 8075 bp in length  
77904 86103 contig of 8200 bp in length  
86204 94612 contig of 8409 bp in length  
94713 103173 contig of 8461 bp in length  
103273 103173 contig of 8461 bp in length  
11019 117944 contig of 7645 bp in length  
111019 117944 contig of 7645 bp in length  
118045 124109 contig of 6065 bp in length  
124210 129179 contig of 4970 bp in length  
129280 136405 contig of 7126 bp in length  
136506 142552 contig of 6047 bp in length  
142553 142552 contig of 6047 bp in length  
142553 146979 contig of 4327 bp in length  
146980 147079 contig of 100 bp in length  
147080 151211 contig of 4132 bp in length  
151212 151311 contig of 100 bp in length  
151312 155117 contig of 3806 bp in length  
155118 155217 contig of 100 bp in length  
155218 158879 contig of 3662 bp in length  
158880 158979 contig of 100 bp in length  
158980 161806 contig of 2827 bp in length  
161807 161906 contig of 100 bp in length  
161907 164032 contig of 2126 bp in length  
164033 164132 contig of 100 bp in length  
164133 165745 contig of 1613 bp in length  
165746 165845 contig of 100 bp in length  
165846 167102 contig of 1257 bp in length  
167103 167202 contig of 100 bp in length  
167203 168212 contig of 1010 bp in length  
168213 168312 contig of 100 bp in length  
168313 170252 contig of 1940 bp in length  
170253 170352 contig of 100 bp in length  
170353 171734 contig of 1382 bp in length  
171735 171835 contig of 100 bp in length  
171835 173810 contig of 1976 bp in length  
173811 173910 contig of 100 bp in length

142653 146979 contig of 4327 bp in length  
147080 151211 contig of 4132 bp in length  
151312 155117 contig of 3806 bp in length  
155218 158879 contig of 3662 bp in length  
158980 161806 contig of 2827 bp in length  
161907 164032 contig of 2126 bp in length  
164133 165745 contig of 1613 bp in length  
165846 167102 contig of 1010 bp in length  
167203 168212 contig of 100 bp in length  
168313 170252 contig of 1940 bp in length  
170253 170352 contig of 100 bp in length  
170353 171734 contig of 1382 bp in length  
171735 171835 contig of 100 bp in length  
171835 173810 contig of 1976 bp in length  
173811 173910 contig of 100 bp in length

Sequence updated (06-Jan-2000)  
Sequence updated (26-May-2000)  
NOTE: This is a 'working draft' sequence. It currently  
consists of 27 contigs. The true order of the pieces  
is not known and their order in this sequence record is  
arbitrary. Gaps between the contigs are represented as  
runs of N, but the exact sizes of the gaps are unknown.  
This record will be updated with the finished sequence  
as soon as it is available and the accession number will  
be preserved.

1 46938 contig of 46938 bp in length  
46939 47038 contig of 100 bp in length  
47039 69628 contig of 22590 bp in length  
69629 69728 contig of 100 bp in length  
69729 77803 contig of 8075 bp in length  
77804 77903 contig of 8200 bp in length  
77904 86103 contig of 8200 bp in length  
86104 86203 contig of 8409 bp in length  
86204 94612 contig of 8409 bp in length  
94613 94712 contig of 100 bp in length  
94713 103173 contig of 8461 bp in length  
103174 103273 contig of 100 bp in length  
103274 103918 contig of 7645 bp in length  
11019 111018 contig of 100 bp in length  
111019 117944 contig of 6926 bp in length  
117945 118044 contig of 100 bp in length  
118045 124109 contig of 6065 bp in length  
124110 124209 contig of 100 bp in length  
124210 129179 contig of 4970 bp in length  
129180 129279 contig of 100 bp in length  
129280 136405 contig of 7126 bp in length  
136406 136505 contig of 100 bp in length  
136506 142552 contig of 6047 bp in length  
142553 142552 contig of 6047 bp in length  
142553 146979 contig of 4327 bp in length  
146980 147079 contig of 100 bp in length  
147080 151211 contig of 4132 bp in length  
151212 151311 contig of 100 bp in length  
151312 155117 contig of 3806 bp in length  
155118 155217 contig of 100 bp in length  
155218 158879 contig of 3662 bp in length  
158880 158979 contig of 100 bp in length  
158980 161806 contig of 2827 bp in length  
161807 161906 contig of 100 bp in length  
161907 164032 contig of 2126 bp in length  
164033 164132 contig of 100 bp in length  
164133 165745 contig of 1613 bp in length  
165746 165845 contig of 100 bp in length  
165846 167102 contig of 1257 bp in length  
167103 167202 contig of 100 bp in length  
167203 168212 contig of 1010 bp in length  
168213 168312 contig of 100 bp in length  
168313 170252 contig of 1940 bp in length  
170253 170352 contig of 100 bp in length  
170353 171734 contig of 1382 bp in length  
171735 171835 contig of 100 bp in length  
171835 173810 contig of 1976 bp in length  
173811 173910 contig of 100 bp in length

FEATURES

\* 173911 175003: contig of 1093 bp in length

\* 175004 175103: gap of 100 bp

\* 175104 176492: contig of 1389 bp in length

\* 176493 176592: gap of 100 bp

\* 176593 177654: contig of 1062 bp in length.

source

1. 177654

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

/chromosome="18"

/map="18p11.3"

/clone="RP11-737C10"

1. 46938

/note="assembly\_fragment"

47039. 69628

/note="assembly\_fragment"

69729. 77803

/note="assembly\_fragment"

77904. 86103

/note="assembly\_fragment"

86204. 94612

/note="assembly\_fragment clone\_end:17 vector\_side:right"

94713. 103173

/note="assembly\_fragment"

103274. 110918

/note="assembly\_fragment"

111019. 117944

/note="assembly\_fragment"

118045. 124109

/note="assembly\_fragment"

124210. 129179

/note="assembly\_fragment"

129280. 136405

/note="assembly\_fragment"

136506. 142552

/note="assembly\_fragment"

142653. 146979

/note="assembly\_fragment"

147080. 151211

/note="assembly\_fragment"

151312. 155117

/note="assembly\_fragment"

155218. 158879

/note="assembly\_fragment"

158980. 161806

/note="assembly\_fragment"

161907. 164032

/note="assembly\_fragment"

164133. 165745

/note="assembly\_fragment"

165846. 167102

/note="assembly\_fragment"

167203. 168212

/note="assembly\_fragment"

168313. 170252

/note="assembly\_fragment"

170353. 171734

/note="assembly\_fragment"

171835. 173810

/note="assembly\_fragment"

173911. 175003

/note="assembly\_fragment"

175104. 176492

Query Match 54.0%; Score 36.2; DB 2; Length 177654;

Best Local Similarity 72.3%; Pred. No. 0.017; Matches 4; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 2 ACAAGCTGGAATTAAGTCTGCCACCAAGTTGTAACCTTTAGGTTTAAGACAGTAC 61

DB 63974 ACAAACTGGAATTAAGTCTGCCACCAAGTTGTAACCTTTAGGTTTAAGACAGTAC 63915

QY 62 AGAGT 66

DB 63914 GAGCT 63910

RESULT 10

AC013412/c

LOCUS

DEFINITION

AC013412

ACCESSION

AC013412.3 GI:8570260

VERSION

KEYWORDS

SOURCE

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 200214)

Suleston, J.B. and Waterston, R.

Toward a complete human genome sequence

Genome Res. 8 (11), 1097-1108 (1998)

99063792

MEDLINE

9847074

2 (bases 1 to 200214)

Bemis, G., Kallio, J., Stoneking, T. and Parker, C.

The sequence of Homo sapiens BAC clone RP11-507A3

Unpublished

3 (bases 1 to 200214)

Waterston, R.H.

Direct Submision

Submitted (09-NOV-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

4 (bases 1 to 200214)

Waterston, R.H.

Direct Submision

Submitted (17-JUN-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

5 (bases 1 to 200214)

Waterston, R.

Direct Submision

Submitted (30-SEP-2000) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

On Jun 17, 2000 this sequence version replaced gi:7631033.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: <http://genome.wustl.edu/gsc>

Contact: [sapiens@wustl.edu](mailto:sapiens@wustl.edu)

Summary Statistics

Center project name: H\_NH0507A03

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The position of this clone was established as part of a collaboration between the Human Chromosome Y Mapping Project (Tomoko Kawaguchi, Helen Skalecky, Laura G. Brown, Steve Rozen, and David C. Page at the Whitehead Institute for Biomedical Research, Cambridge MA) and the Washington University Genome Sequencing Center, St. Louis MO.



RESULT 11  
D83730 1935 bp DNA linear PRI 10-JUL-2001  
LOCUS Homo sapiens AMGX gene for amelogenin, intron 2.  
DEFINITION D83730.1 GI:5263179  
ACCESSION D83730.1 GI:5263179  
VERSION  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
AUTHORS Sekiguchi, H., Minaguchi, K., Machida, Y. and Yakushiji, M.  
TITLE PCR detection of the human amelogenin gene and its application to the diagnosis of amelogenesis imperfecta  
JOURNAL Bull. Tokyo Dent. Coll. 39 (4), 275-285 (1998)  
MEDLINE 99234629  
PUBMED 10218009

REFERENCE  
AUTHORS Sekiguchi, H.  
TITLE Direct Submission  
JOURNAL Submitted (29-FEB-1996) Hiroshi Sekiguchi, Tokyo Dental College, Pediatric Dentistry, 1-2-2 Masago, Minama-ku, Chiba 261, Japan (Tel:043-270-3945, Fax:043-279-2052)  
Location/Qualifiers  
1. .1935  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="X"

FEATURES  
source  
1. .1935  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="X"

gene  
1. .1935  
/gene="AMGX"  
1. .1935  
/gene="AMGX"  
/note="amelogenin  
this nucleotide sequence shows an intron between exon2 and exon3 within human amelogenin gene on the X chromosome"  
/number=2

ORIGIN  
Query Match 52.8%; Score 35.4; DB 9; Length 1935;  
Best Local Similarity 80.6%; Pred. No. 0.032; Mismatches 54; Conservative 0; Mismatches 11; Indels 2; Gaps 1;  
Matches 54; Conservative 0; Mismatches 11; Indels 2; Gaps 1;

QY 1 CACAAAGCTTGAATAAAGTCTGCCCCACAGTTGTAATTAGGTTTAAAGACAGTA 60  
|||||  
40 CACAAAGCTTGAATAAAGTCTGCCCC-ATAGTTGTAATAATTAGGTTTAAAGACAGTA 97  
|||||

DB 61 CAAGATC 67  
|||||  
98 TGAGATC 104

RESULT 12  
AB091794 5591 bp DNA linear MAM 02-MAY-2003  
LOCUS AB091794  
DEFINITION Equus caballus AMELY gene for amelogenin, partial cds.  
ACCESSION AB091794  
VERSION  
KEYWORDS  
SOURCE Equus caballus (horse)  
ORGANISM Equus caballus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Perissodactyla; Equidae; Equus.

REFERENCE  
AUTHORS Iwase, M., Satta, Y., Hirai, Y., Hirai, H., Imai, H. and Takahata, N.  
TITLE From the Cover: The amelogenin loci span an ancient pseudautosomal boundary in diverse mammalian species  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 100 (9), 5258-5263 (2003)  
MEDLINE 12672962  
PUBMED 12672962

REFERENCE  
AUTHORS Iwase, M., Satta, Y. and Takahata, N.  
TITLE Direct Submission  
JOURNAL Submitted (19-SEP-2002) Mineyo Iwase, Graduate University for Advanced Studies (Sokendai), Department of Biosystems Science; Shonan Kokusaijima, Hayama, Kanagawa 240-0193, Japan (E-mail: iwase@koryu.w01.soken.ac.jp, Tel: 81-468-58-1571, Fax: 81-468-58-1544)  
Location/Qualifiers  
1. .5591  
/organism="Equus caballus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9796"  
/sex="male"

FEATURES  
source  
1. .5591  
/organism="Equus caballus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9796"  
/sex="male"

gene  
130. .5591  
/gene="AMELY"  
join(130. .180, 1493. .1494)  
/gene="AMELY"  
join(1495. .1548, 3470. .3517, 4907. .4915, 5235. .5591)  
/gene="AMELY"  
/codon\_start=1  
/product="amelogenin"  
/protein\_id="BAC66113.1"  
/db\_xref="GI:29126041"  
/translation="MGTWLFACTVGTATAMPPLRPHRPHAYINSEVLTPTSYGYE  
PMSGLHQIIPVLTQONPSNHLQPHHHPVPAQGVSVSOQPAIPLEGCHSMIPSQ  
HHQRLPRLPVQPPQPPHQPRIQPPHPIQPLPQAPLPPIPLAASGRPAS"

ORIGIN  
Query Match 52.2%; Score 35; DB 4; Length 5591;  
Best Local Similarity 70.1%; Pred. No. 0.046; Mismatches 47; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 1 CACAAAGCTTGAATAAAGTCTGCCCCACAGTTGTAATTAGGTTTAAAGACAGTA 60  
|||||  
DB 1574 CACAAAGCTTGAATAAAGTCTGCCCCCTCTTGTGTGTAATTTTGGGTTTAAATCAATA 1633  
|||||

QY 61 CAAGATC 67  
|||||  
DB 1634 CAAGATC 1640

RESULT 13  
AB091783 6442 bp DNA linear PRI 02-MAY-2003  
LOCUS AB091783  
DEFINITION Saimiri sciureus AMELX gene for amelogenin, partial cds.  
ACCESSION AB091783  
VERSION  
KEYWORDS  
SOURCE Saimiri sciureus (common squirrel monkey)  
ORGANISM Saimiri sciureus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Cebinae; Saimiri.

REFERENCE  
AUTHORS Iwase, M., Satta, Y., Hirai, Y., Hirai, H., Imai, H. and Takahata, N.  
TITLE From the Cover: The amelogenin loci span an ancient pseudautosomal boundary in diverse mammalian species  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 100 (9), 5258-5263 (2003)  
MEDLINE 12672962  
PUBMED 12672962

REFERENCE  
AUTHORS Iwase, M., Satta, Y. and Takahata, N.  
TITLE Direct Submission  
JOURNAL Submitted (19-SEP-2002) Mineyo Iwase, Graduate University for Advanced Studies (Sokendai), Department of Biosystems Science; Shonan Kokusaijima, Hayama, Kanagawa 240-0193, Japan (E-mail: iwase@koryu.w01.soken.ac.jp, Tel: 81-468-58-1571, Fax: 81-468-58-1544)  
Location/Qualifiers  
1. .6442  
/organism="Saimiri sciureus"  
/mol\_type="genomic DNA"

```

/db_xref="taxon:9521"
/gen="male"
/gene="AMELX"
/join(936..6442
/gene="AMELX"
/join(936..991,2298..2309)
/gene="AMELX"
/join(2310..2363,4289..4336,5592..5633,5725..5769,
6041..56442)
/gene="AMELX"
/codon_start=1
/product="amelogenin"
/protein_id="BAC66103.1"
/db_xref="GI:29126020"
/translation="MGTWILFACLTGAAPAMP.LPHRGHGTYNFSYENSHSQAINID
KRGVLTPLKMYOSMIRPVRPSYSGYRPMGMLHQIIPVLSQHPPLTHLQSHHLPV
VPAQDPVFPQDPMPVFGQHSMTPTQHQNPPLPAAQDPQFQTVQPPQHPQPPQPP
VHMQPLPQPPPLPMPFPMQPLPPLPDLLEAWPAT"

ORIGIN

Query Match          51.6%; Score 34.8; DB 9; Length 6442;
Best Local Similarity 80.3%; Pred. No. 0.054;
Matches 53; Conservative 0; Mismatches 12; Indels 1; Gaps 1;

QY 1 CACAACTTGAATTAAGCTGCCCCACCACTTGCTTAAGGTTTAAGACAGTA 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2401 CAAATATTAAGATTAAGATCTGCCCCA-TGGTTGTTGAAGTTTAAGACAGTA 2459
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 61 CAAGAT 66
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2460 TGAGAT 2465

RESULT 14
AB091782          7163 bp      DNA      linear      PRI 02-MAY-2003
LOCUS             Pan troglodytes AMELY gene for amelogenin, partial cds.
DEFINITION        AB091782
ACCESSION          AB091782
VERSION            AB091782.1 GI:29126017
KEYWORDS
SOURCE             Pan troglodytes (chimpanzee)
ORGANISM            Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.
REFERENCE          1 Iwase,M., Satta,Y., Hirai,Y., Hirai,H., Imai,H. and Takahata,N.
AUTHORS            From the Cover: The amelogenin loci span an ancient pseudautosomal
TITLE              boundary in diverse mammalian species
JOURNAL            Proc. Natl. Acad. Sci. U.S.A. 100 (9), 5258-5263 (2003)
MEDLINE            22608569
PUBMED             12672962
REFERENCE          2 (bases 1 to 7163)
AUTHORS            Iwase,M., Satta,Y. and Takahata,N.
TITLE              Submitted (19-SEP-2002) Mineyo Iwase, Graduate University for
JOURNAL            Advanced Studies(Soken-dai), Department of Biosystems Science;
                    Shonan Kokusai-mura, Hayama, Kanagawa 240-0193, Japan
                    (E-mail:iwase@koryu.w01.soken.ac.jp, Tel:81-468-58-1571,
                    Fax:81-468-58-1544)
FEATURES
    source          1. 7163
                    Location/Qualifiers
                    1. 7163
                    /organism="Pan troglodytes"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:9598"
                    /sex="male"
                    1186..7163
                    /gene="AMELY"
                    join(1186..1241,2593..2604)
                    /gene="AMELY"
                    join(2605..2658,5151..5198,6297..6338,6430..6474,
                    6744..67163)
                    /gene="AMELY"
                    /codon_start=1
                    /product="amelogenin"

```

```

/protein_id="BAC66102.1"
/db_xref="GI:29126018"
/translation="MGTWILFACLTGAAPAMP.LPHRGHGTYNFSYENSHSQAINID
KIALVLTPLKMYOSMIRPVSYSYRPMGMLHQIIPVLSQHPPLTHLQSHHLPV
VPAQDPVFPQDPMPVFGQHSMTPTQHQNPPLPAAQDPQFQTVQPPQHPQPPQPP
VQPMQPLPQPPPLPMPFPMRPLPPLPDLLEAWPATDKTKKE"

ORIGIN

Query Match          51.6%; Score 34.6; DB 9; Length 7163;
Best Local Similarity 70.8%; Pred. No. 0.064;
Matches 46; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 2 ACAAGCTGAATTAAGCTGCCACCACTTGCTTAAGGTTTAAGACAGTAC 61
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3732 ACAATGTGAATTAAGATTTTCTCCCACTGGAACCTTAAGTTTAACAGTTT 3673
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 62 AAGAT 66
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3672 GAGCT 3668

RESULT 15
BS000568/c        38765 bp      DNA      linear      PRI 17-MAR-2004
LOCUS             Pan troglodytes chromosome Y clone:PTFY-001K12, complete sequences.
DEFINITION        BS000568
ACCESSION          BS000568
VERSION            BS000568.1 GI:45504168
KEYWORDS
SOURCE             Pan troglodytes (chimpanzee)
ORGANISM            Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.
REFERENCE          1 The Chimpanzee Chromosome Y Sequencing Consortium.
AUTHORS            DNA sequence of chimpanzee chromosome Y and its evolutionary
TITLE              implications
JOURNAL            2 (bases 1 to 38765)
AUTHORS            Hattori,M., Toyoda,A., Noguchi,H., Taylor,T.D., Kuroki,Y.,
                    Fujiyama,A. and Sakaki,Y.
TITLE              Direct Submision
JOURNAL            Submitted (15-MAR-2004) Masahira Hattori, The Institute of Physical
                    and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
                    1-7-22 Suenho-chou,Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan
                    (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.jp/,
                    Tel:81-45-503-9111, Fax:81-45-503-9170)
FEATURES
    source          1. 38765
                    Location/Qualifiers
                    1. 38765
                    /organism="Pan troglodytes"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:9598"
                    /chromosome="Y"
                    /clone="PTFY-001K12"

ORIGIN

Query Match          51.6%; Score 34.6; DB 9; Length 38765;
Best Local Similarity 70.8%; Pred. No. 0.066;
Matches 46; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 2 ACAAGCTGAATTAAGCTGCCACCACTTGCTTAAGGTTTAAGACAGTAC 61
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3732 ACAATGTGAATTAAGATTTTCTCCCACTGGAACCTTAAGTTTAACAGTTT 3673
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 62 AAGAT 66
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3672 GAGCT 3668

Search completed: February 8, 2005, 16:36:59
Job time : 377.57 secs

```

**This Page Blank (uspto)**



CC sarcoma, possibly via gene therapy or vaccine production. The nucleic  
CC acid sequences may be useful in diagnostic and screening applications.  
CC The current sequence is that of a human soft tissue sarcoma-upregulated  
CC DNA of the invention. The current sequence is not shown within the  
CC specification per se but was submitted in CD format by the inventor.

XX Sequence 259202 BP; 81699 A; 51421 C; 49221 G; 76861 T; 0 U; 0 Other;

Query Match 50.4%; Score 33.9; DB 12; Length 259202;

Best Local Similarity 79.1%; Pred. No. 0.045; Mismatches 12; Indels 2; Gaps 1;

Matches 53; Conservative 0; Mismatches 12; Indels 2; Gaps 1;

QY 1 CACAGCTTGGAATAAAGTCTGCCCCAGCTTGTAATTAGGTTTAAAGACAGTA 60  
DB 69966 CACAACTTGACATATAAATCTGCTC-ATAGTTGTGAATTAGGTTTAAAGACGTA 70023

QY 61 CAGATC 67

DB 70024 TGAGATC 70030

RESULT 2

ABV98228/c

ID ABV98228 standard; cDNA; 559 BP.

AC ABV98228;

DT 14-JAN-2003 (first entry)

DE Human pancreatic cancer expressed cDNA SEQ ID NO 3636.

KM Human; pancreas; cancer; gene therapy; vaccine; immunostimulant;  
KW cytosolic; tumour; gene; ss.

OS Homo sapiens.

PN WO200260317-A2.

PD 08-ANG-2002.

PF 30-JAN-2002; 2002WO-US002781.

PR 30-JAN-2001; 2001US-0265305P.

PR 31-JAN-2001; 2001US-0265682P.

PR 09-FEB-2001; 2001US-0267588P.

PR 21-MAR-2001; 2001US-0278651P.

PR 28-APR-2001; 2001US-0287112P.

PR 16-MAY-2001; 2001US-0291631P.

PR 12-JUL-2001; 2001US-0305484P.

PR 20-AUG-2001; 2001US-0313999P.

PR 27-NOV-2001; 2001US-0333626P.

(CORI-) CORIXA CORP.

Benson DR, Kalos MD, Lodes MJ, Persing DH, Heppler WT, Jiang Y;

WPI; 2002-627435/67.

PT New isolated polynucleotide and pancreatic tumor polypeptides, useful for  
PT diagnosing, preventing and/or treating cancer, particularly pancreatic  
PT cancer.

PS Claim 1; SEQ ID NO 3636; 300bp + Sequence listing; English.

CC The invention relates to an isolated polynucleotide (I) comprising: (a)  
CC any of a group of over 4000 nucleotide sequences (ABV94628-ABV99145); (b)  
CC complements of (a); (c) sequences consisting of at least 20 contiguous  
CC residues of (a); (d) sequences that hybridize to (a) under moderately  
CC stringent conditions; (e) sequences having at least 75% or 50% identity  
CC to (a); or (f) degenerate variants of (a). Polypeptides (ABP68596-  
CC ABP68637) encoded by (I) and oligonucleotide can be used to detect cancer  
CC in a patient and compositions comprising polypeptides, polynucleotides,  
CC antibodies, fusion proteins, T cell populations and antigen presenting

CC cells expressing the polypeptide are useful in treating pancreatic cancer  
CC and stimulating an immune response. The polynucleotides can be used as  
CC probes or primers for nucleic acid hybridization, in the design and  
CC preparation of ribozyme molecules for inhibiting expression of the tumour  
CC polypeptide and proteins in the tumour cells, in vaccines and for gene  
CC therapy. Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp://wipo.int/pub/published\_pct\_sequences

XX Sequence 559 BP; 193 A; 84 C; 111 G; 171 T; 0 U; 0 Other;

Query Match 39.7%; Score 26.6; DB 6; Length 559;

Best Local Similarity 63.1%; Pred. No. 5.7; Mismatches 24; Indels 0; Gaps 0;

Matches 41; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 3 CAGCTTGGAATAAAGTCTGCCCCAGCTTGTAATTAGGTTTAAAGACAGTACA 62  
DB 293 CAAATATGATATAAATGTTTGTGCCCCAGCTTGTAATTCTTAAGATTAAGATACAA 234

QY 63 AGATC 67

DB 233 TGATC 229

RESULT 3

ID ABQ60609 standard; cDNA; 614 BP.

AC ABQ60609;

DT 02-AUG-2002 (first entry)

DE Human colon cancer related nucleotide sequence SEQ ID NO:4304.

KM Human; colon cancer; cancer; tissue profiling; forensic; mapping;  
KW genetic analysis; diagnostic; antisense therapy; gene; ss.

OS Homo sapiens.

PN WO200229086-A2.

PD 11-APR-2002.

PR 02-OCT-2001; 2001WO-US030732.

PR 02-OCT-2000; 2000US-0237271P.

(FARB) BAYER CORP.

Burgess C, Aistle JH, Carroll E, Catino TV, Dwivedi P, Molino GA;

Thiagalingam A, Lewis ME;

WPI; 2002-426115/45.

PT New isolated nucleic acid that is differentially expressed in cancer  
PT tissues useful for determining the presence of colon cancer in a cell or  
PT tissue type, and in antisense therapy.

PS Claim 1; Fig 1; 796bp; English.

CC ABQ56306 to ABQ60787 represent isolated nucleic acids (I) differentially  
CC expressed in cancer tissues. ABQ78993 to ABQ79004 represent proteins  
CC encoded by the ABQ60776 to ABQ60787 nucleic acid sequences. (I) can be  
CC used in antisense therapy. An antibody immunoreactive with a polypeptide  
CC encoded by (I) is useful for detecting cancer in a patient sample, and  
CC for detecting the presence or absence of a polynucleotide encoded by a  
CC nucleic acid which hybridizes to (I) in a cell. A probe/primer derived  
CC from (I) can be used for determining the presence of a nucleic acid which  
CC hybridizes to (I), and for determining the phenotype of cells in a sample  
CC of cells from a patient. (I) is useful for determining the presence of  
CC colon cancer in a cell or tissue type, for determining the presence or  
CC state of other type of cancer, in antisense therapy, to generate  
CC microarrays on a solid surface, to identify a chromosome on which the



CC corresponding gene residues, and in tissue profiling, forensics, genetic  
CC analysis, mapping and diagnostic applications. (I) can be used to raise  
CC antibodies, and to screen for peptide analogues and antagonists  
XX

SQ Sequence 614 BP, 186 A; 126 C; 81 G; 216 T; 0 U; 5 Other;

Query Match 39.7%; Score 26.6; DB 6; Length 614;  
Best Local Similarity 63.1%; Pred. No. 5.8; 24; Indels 0; Gaps 0;  
Matches 41; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 3 CAGCTTGGAATTAATTAAGTCTGCCCCCAGCTTGTTAGGCTTTAAGACAGTACA 62  
DB 261 CAAATATGATTAATTAATTAAGTCTGCCCCCAGCTTGTTATCTCTAAGATTAAGACTATCA 320

QY 63 AGATC 67  
DB 321 TGATC 325

## RESULT 4

ADP12803/c  
ID ADP12803 standard; DNA; 2176 BP.

AC ADP12803;

DT 12-AUG-2004 (first entry)

DE Reference mRNA sequence #17.

XX transplant rejection; immune system; rheumatoid arthritis; lupus;

KW inflammatory bowel disease; multiple sclerosis; HIV; AIDS; ss.

XX Homo sapiens.

PN WO2004042346-A2.

PD 21-MAY-2004.

PF 24-APR-2003; 2003WO-US012946.

PR 24-APR-2002; 2002US-00131831.

PR 20-DEC-2002; 2002US-00325899.

XX (EXPR-) EXPRESSION DIAGNOSTICS INC.

PI Woljgemuth J, Fry K, Woodward R, Ly N, Prentice J, Morris M;

PI Rosenberg S;

DR MPI; 2004-400724/37.

XX Diagnosing or monitoring transplant rejection, e.g. heart, kidney, liver,  
PT pancreas, pancreatic islet, lung, bone marrow or stem cell transplant  
PT rejection, in an individual, comprises detecting the expression level of  
PT the genes.

PS Claim 65; SEQ ID NO 2812; 1762bp; English.

XX The present invention relates to diagnosing or monitoring transplant  
CC rejection, e.g. cardiac or kidney transplant rejection, in an individual  
CC comprising detecting the expression level of one or more genes. The  
CC methods, system and kits are useful in diagnosing or monitoring  
CC transplant rejection, e.g. heart, kidney, liver, pancreas, pancreatic  
CC islet, lung, bone marrow or stem cell transplant rejection,  
CC xenotransplant rejection or mechanical organ replacement rejection, in an  
CC individual. The method is also useful in assessing the immune status of  
CC an individual. The methods are also useful in diagnosing and monitoring  
CC diseases that involve the immune system, e.g. rheumatoid arthritis,  
CC lupus, inflammatory bowel diseases, multiple sclerosis, HIV/AIDS or  
CC viral, bacterial or fungal infection. The present sequence represents a  
CC reference mRNA sequence of the invention which show altered expression in  
CC renal transplantation and expression.

XX Sequence 2176 BP, 770 A; 309 C; 425 G; 672 T; 0 U; 0 Other;

Query Match 39.7%; Score 26.6; DB 12; Length 2176;  
Best Local Similarity 63.1%; Pred. No. 7.9; 24; Indels 0; Gaps 0;  
Matches 41; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 3 CAGCTTGGAATTAATTAAGTCTGCCCCCAGCTTGTTAGGCTTTAAGACAGTACA 62  
DB 1471 CAAATATGATTAATTAATTAAGTCTGCCCCCAGCTTGTTATCTCTAAGATTAAGACTATCA 1412

QY 63 AGATC 67  
DB 1411 TGATC 1407

## RESULT 5

ADQ24473/c  
ID ADQ24473 standard; DNA; 2516 BP.

AC ADQ24473;

DT 26-AUG-2004 (first entry)

DE Human soft tissue sarcoma-upregulated DNA - SEQ ID 7293.

XX soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;  
KW ds.

XX Homo sapiens.

PN WO2004048938-A2.

PD 10-JUN-2004.

PF 26-NOV-2003; 2003WO-US038193.

PR 26-NOV-2002; 2002US-0429739P.

XX (PROT-) PROTEIN DESIGN LABS INC.

PI Aziz N, Ginsburg WM, Zlotnik A;

DR MPI; 2004-441208/41.

XX Early detection of soft tissue sarcoma comprises determining expression  
PT of a gene in a first soft tissue sample and a normal soft tissue sample  
PT and comparing the gene expression, also useful in treating soft tissue  
PT sarcoma.

PS Example 2; SEQ ID NO 7293; 210bp; English.

XX The invention relates to a novel method for detecting soft tissue sarcoma  
CC which comprises obtaining a first soft tissue sample from an individual  
CC and a normal soft tissue sample from the same or different individual,  
CC determining the expression of a gene in both samples and comparing the  
CC expression of the gene in both soft tissue samples, where a higher level  
CC of protein expression in the first soft tissue sample indicates the  
CC presence of soft tissue sarcoma. The method of the invention has  
CC cytostatic applications and may be useful for detecting soft tissue  
CC sarcoma, possibly via gene therapy or vaccine production. The nucleic  
CC acid sequences may be useful in diagnostic and screening applications.  
CC The current sequence is that of a human soft tissue sarcoma-upregulated  
CC DNA of the invention. The current sequence is not shown within the  
CC specification per se but was submitted in CD format by the inventor.

XX Sequence 2516 BP; 862 A; 376 C; 484 G; 779 T; 0 U; 15 Other;

Query Match 39.7%; Score 26.6; DB 12; Length 2516;  
Best Local Similarity 63.1%; Pred. No. 8.1; 24; Indels 0; Gaps 0;  
Matches 41; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 3 CAGCTTGGAATTAATTAAGTCTGCCCCCAGCTTGTTAGGCTTTAAGACAGTACA 62  
DB 1665 CAAATATGATTAATTAATTAAGTCTGCCCCCAGCTTGTTATCTCTAAGATTAAGACTATCA 1606

OY 63 AGATC 67  
DB 1605 TGATC 1601

## RESULT 6

ADQ22895/c  
ID ADQ22895 standard; DNA; 2516 BP.

XX ADQ22895;

XX 26-AUG-2004 (first entry)

XX Human soft tissue sarcoma-upregulated DNA - SEQ ID 5715.

XX soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;  
KM db.

XX Homo sapiens.

XX WO2004048938-A2.

XX 10-JUN-2004.

XX 26-NOV-2003; 2003WO-US038193.

XX 26-NOV-2002; 2002US-0429739P.

XX (PROT-) PROTEIN DESIGN LABS INC.

XX Aziz N, Ginsburg WM, Zlotnik A;

XX WPI; 2004-441208/41.

XX Early detection of soft tissue sarcoma comprises determining expression  
PT of a gene in a first soft tissue sample and a normal soft tissue sample  
PT and comparing the gene expression, also useful in treating soft tissue  
PT sarcoma.

XX Example 2; SEQ ID NO 5715; 210pp; English.

XX The invention relates to a novel method for detecting soft tissue sarcoma  
CC which comprises obtaining a first soft tissue sample from an individual  
CC and a normal soft tissue sample from the same or different individual,  
CC determining the expression of a gene in both samples and comparing the  
CC expression of the gene in both soft tissue samples, where a higher level  
CC of protein expression in the first soft tissue sample indicates the  
CC presence of soft tissue sarcoma. The method of the invention has  
CC cytostatic applications and may be useful for detecting soft tissue  
CC sarcoma, possibly via gene therapy or vaccine production. The nucleic  
CC acid sequences may be useful in diagnostic and screening applications.  
CC The current sequence is that of a human soft tissue sarcoma-upregulated  
CC DNA of the invention. The current sequence is not shown within the  
CC specification per se but was submitted in CD format by the inventor.

XX Sequence 2516 BP; 862 A; 376 C; 484 G; 779 T; 0 U; 15 Other;

XX Query Match 39.7%; Score 26.6; DB 12; Length 2516;

XX Best Local Similarity 63.1%; Pred. No. 8.1; Mismatches 24; Indels 0; Gaps 0;

XX Matches 41; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

OY 3 CAAGCTTGAATAAAGTCTGCCCCACAGTTGGTAACCTTAGGTTTAAGACAGTACA 62

DB 1665 CAATATATATATATATATATATGTTGTGCCCCAGTTTATCTTAAGATAGATATCA 1606

OY 63 AGATC 67

## RESULT 7

AD161706/c

ID AD161706 standard; cDNA; 9262 BP.

XX AD161706;

XX 22-APR-2004 (first entry)

XX Human cDNA downregulated in Alzheimer's disease, INCYTE 24655.92.

XX Human; ss; Alzheimer's disease; differential display; neuroprotective;  
KM brain disorder.

XX Homo sapiens.

XX US6682888-B1.

XX 27-JAN-2004.

XX 05-MAY-2000; 2000US-00566921.

XX 05-MAY-2000; 2000US-00566921.

XX (INCY-) INCYTE CORP.

XX Loring JF, Tingley DW, Edwards CM;

XX WPI; 2004-118572/12.

XX New composition comprising cDNAs that are differentially expressed in  
PT brain disorders, useful for diagnosing or treating Alzheimer's disease.

XX Claim 1; SEQ ID NO 74; 223pp; English.

XX The invention relates to a new composition comprising AD161633-  
CC AD161706 and their complements that are cDNAs differentially expressed in  
CC brain disorders. Also included are a high throughput method for detecting  
CC differential expression of one or more cDNAs in a sample containing  
CC nucleic acids and a high throughput method for screening a library of  
CC molecules or compounds to identify a ligand that specifically binds a  
CC cDNA. The expression of the each of the cDNAs is downregulated at least  
CC two-fold in the brain of the subjects with Alzheimer's disease (AD161727-  
CC AD161727) or upregulated at least two fold in Alzheimer's disease  
CC (AD161728-AD161770). The composition is useful for diagnosing or treating  
CC Alzheimer's disease. The present sequence is a cDNA downregulated at  
CC least two-fold in the brain of the subjects with Alzheimer's disease.

XX Sequence 9262 BP; 2795 A; 1687 C; 1820 G; 2951 T; 0 U; 9 Other;

XX Query Match 39.7%; Score 26.6; DB 12; Length 9262;

XX Best Local Similarity 63.1%; Pred. No. 11; Mismatches 24; Indels 0; Gaps 0;

XX Matches 41; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

OY 3 CAAGCTTGAATAAAGTCTGCCCCACAGTTGGTAACCTTAGGTTTAAGACAGTACA 62

DB 5902 CAATATATATATATATATATGTTGTGCCCCAGTTTATCTTAAGATAGATATCA 5843

OY 63 AGATC 67

DB 5842 TGATC 5838

RESULT 8

ADJ56289/c

ADJ56289 standard; cDNA; 9277 BP.

ADJ56289;

06-MAY-2004 (first entry)

Human cDNA differentially expressed in MYCN activated cells SeqID 95.

human; differential expression; transactivator; proto-oncogene;  
KM neuroblastoma; small cell lung cancer; cytostatic; gene therapy; ss;  
KM MYCN activated cell.

```
XX Homo sapiens.
OS
XX US2003119009-A1.
XX
XX 26-JUN-2003.
XX
XX 25-FEB-2002; 2002US-00084817.
XX
XX 23-FEB-2001; 2001US-0270784P.
XX
XX (STUA/) STUART S G.
XX (NUCH/) NUCHTEN J G.
XX (PLON/) PLON S E.
XX (SHOH/) SHOHEI J M.
XX
XX Stuart SG, Nuchtern JG, Plon SE, Shohe J M;
XX
XX MPI; 2003-635698/60.
XX
XX New genes regulated by MYCN activation, useful in gene therapy,
XX particularly for treating a subject with e.g. neuroblastoma or other
XX cancers, or for diagnosing, staging or monitoring the treatment of the
XX cancer.
XX
XX Claim 1, SEQ ID NO 95; 27bp; English.
XX
XX This invention relates to novel isolated cDNAs that are differentially
XX expressed in MYCN activated cells. Specifically, it refers to
XX polynucleotide sequences that exhibit differential expression patterns in
XX cells activated by the transactivator MYCN, where MYCN is a proto-
XX oncogene that is amplified in neuroblastoma cells and is common in small
XX cell lung cancers. The present invention describes these cDNA molecules
XX as useful for in hybridization assays to detect expression of nucleic
XX acids (or complementary nucleic acids) in a present in a given sample, as
XX well as for screening assays by identifying molecules or compounds that
XX specifically bind the cDNA as a ligand and modulate function or activity.
XX Accordingly, these compositions exhibit cytostatic activity and can also
XX be used for gene therapy purposes. This polynucleotide sequence is a cDNA
XX that is differentially expressed in MYCN activated cells, given in an
XX exemplification of the invention. NOTE: This sequence does not appear in
XX the printed specification but has been obtained in electronic format from
XX the US Patent Office at:
XX ftp.segdata.neptco.gov/sequence.html?docid=20030119009.
XX
XX Sequence 9277 BP; 2805 A; 1687 C; 1820 G; 2955 T; 0 U; 10 Other;
XX
XX Query Match 39.7%; Score 26.6; DB 10; Length 9277;
XX Best Local Similarity 63.1%; Pred. No. 11;
XX Matches 41; Conservative 0; Mismatches 24; Indels 0; Gaps 0;
XX
XX QY 3 CAAGCTTGGAAATAAAGTCTGCCACAGTTGTAATTAGGGTTTAAGACAGTACA 62
XX DB 5902 CAATATGATTAATTAATGTTGTGCCCCGAGTTGTTATCTTAAGATTAATCA 5843
XX
XX QY 63 AGATC 67
XX DB 5842 TGATC 5838
XX
XX RESULT 9
XX ADC86176
XX ID ADC86176 standard; DNA; 47804 BP.
XX
XX ADC86176;
XX
XX 01-JAN-2004 (first entry)
XX
XX Human GPCR gene SEQ ID NO:629.
XX
XX de; gene; human; GPCR;
XX KW guanosine triphosphate-binding protein coupled receptor; gene therapy.
XX
```

```
OS Homo sapiens.
XX
XX EPI270724-A2.
XX
XX 02-JAN-2003.
XX
XX 18-JUN-2002; 2002EP-00013517.
XX
XX 18-JUN-2001; 2001JP-00246789.
XX
XX (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
XX (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATOR.
XX
XX Suwa M, Arai K, Akiyama Y, Aburatani H;
XX
XX MPI; 2003-315783/31.
XX
XX P-PSDB; ADC86177.
XX
XX New polynucleotide, useful for preparing a composition for treating a
XX patient in need of increased or suppressed activity or expression of the
XX guanosine triphosphate-binding protein coupled receptor.
XX
XX Claim 1, SEQ ID NO 629; 28bp; English.
XX
XX The invention relates to a novel polynucleotide encoding a guanosine
XX triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
XX the invention may have a use in gene therapy. The polynucleotide and
XX polypeptide are useful for preparing a composition for treating a patient
XX in need of increased or suppressed activity or expression of the
XX guanosine triphosphate-binding protein coupled receptor. The
XX polynucleotide sequences shown in ADC85548-ADC87616 encode GPCR's of the
XX invention.
XX
XX Sequence 47804 BP; 13080 A; 10916 C; 11116 G; 12192 T; 0 U; 500 Other;
XX
XX Query Match 39.1%; Score 26.2; DB 10; Length 47804;
XX Best Local Similarity 63.5%; Pred. No. 23;
XX Matches 40; Conservative 0; Mismatches 23; Indels 0; Gaps 0;
XX
XX QY 4 AAGCTTGAATTAATTAAGTCTGCCACAGTTGTAATTAGGGTTTAAGACAGTACA 63
XX DB 38451 AAATTAAGTAATTAAGAGCCCCCCCCATGCAAGTGAATTATGATTAAGATCAAT 38510
XX
XX QY 64 GAT 66
XX DB 38511 GAT 38513
XX
XX RESULT 10
XX ACH13083/C
XX ID ACH13083 standard; cDNA; 455 BP.
XX
XX ACH13083;
XX
XX 13-OCT-2003 (first entry)
XX
XX Human adult brain cDNA #295.
XX
XX Human genome mapping; biodiversity; genetic disorder.
XX
XX Homo sapiens.
XX
XX US2003073623-A1.
XX
XX 17-APR-2003.
XX
XX 30-UTL-2001; 2001US-00918995.
XX
XX 30-UTL-2001; 2001US-00918995.
XX
XX (DRMA/) DRMANAC R T.
XX (LABA/) LABAT I.
XX
```

PA	(STAC//) STACHE-CRAIN B.
PA	(DICK//) DICKSON M C.
PA	(JONE//) JONES L W.
PI	Dymanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;
DR	WPI, 2003-615964/58.
XX	
XX	New polynucleotide sequences obtained from various cDNA libraries, useful
PT	as hybridization probes, as oligomers for PCR, for chromosome and gene
PT	mapping, in the recombinant production of protein, or in generating
PT	antisense DNA or RNA.
XX	
PS	Claim 1; SEQ ID NO 295; 44pp; English.
CC	
CC	The invention relates to an isolated polynucleotide comprising any one of
CC	38043 cDNA sequences, appearing as ACN12789-ACH50831, whose sequence was
CC	determined by the technique of SBH (sequencing by hybridisation). Also
CC	included is a purified polypeptide comprising a sequence corresponding to
CC	a reading frame of the novel polynucleotide. The nucleic acid sequences
CC	are useful in diagnostics as expressed sequence tags (EST) for
CC	identifying expressed genes or for physical mapping of the human genome,
CC	in forensics, in assessing biodiversity, or in identifying mutations
CC	responsible for genetic disorders and other traits. The nucleotide
CC	sequences are also useful as hybridisation probes, as oligomers for PCR,
CC	for chromosome and gene mapping, in the recombinant production of
CC	protein, or in generating antibodies specific for it. The purified polypeptide
CC	is useful for generating antibodies specific for it. The present sequence
CC	is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data
CC	for this patent did not form part of the printed specification, but was
CC	obtained in electronic format directly from USPRO at
CC	seqdata.uspro.gov/sequence.html?docid=20030073623
XX	
SO	Sequence 455 BP; 92 A; 103 C; 165 G; 92 T; 0 U; 3 Other;
Qy	Query Match                      38.2%; Score 25.6; DB 9; Length 455;
Db	Best Local Similarity    66.1%; Pred. No. 13;
	Matches 37; Conservative    0; Mismatches    19; Indels     0; Gaps        0
OY	3 CAAGCTTGGAAATPAAAAGTCGCCCCCACCAGTTGGTAATTAGGCGTTAAGAAG 58
DB	160 CCAGCCTTGACCTGCGCTTCTGCTGCCTCACCAAGTTGGTACTTTTGTTTGAAGAG 105
XX	
RESULT 11	
ID	ADA53256 standard; cDNA, 2100 BP.
XX	
AC	ADA53256;
XX	
DT	20-NOV-2003 (first entry)
XX	
DE	Human coding sequence, SEQ ID 824.
XX	
KW	Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;
KW	Gene Therapy; human; secretory protein; membrane proteins; cancer;
KW	Inflammatory disease; osteoporosis; neurological disease; gene; ss.
OS	Homo sapiens.
XX	
PN	EP1293569-A2.
PD	19-MAR-2003.
XX	
PF	21-MAR-2002; 2002EP-0006586.
PR	14-SEP-2001; 2001JP-00328381.
PR	24-JAN-2002; 2002US-0350435P.
XX	
PA	(HELI-) HELIX RES INST.
PA	(REAS-) RES ASSOC BIOTECHNOLOGY.
XX	
II	Iteogi T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;

PI	Yamamoto J., Isono Y., Hio Y., Otsuka K., Nagai K., Irie R., Tamechika I.,
PI	Seki N., Yoshikawa T., Otsuka M., Nagahari K., Masuno Y.,
XX	
DR	WPI, 2003-395539/38.
XX	
XX	
PT	New polynucleotides encoding full-length polypeptides, e.g. secretory
PT	and/or membrane proteins, useful for developing medicines for diseases in
PT	which the gene is involved, or as target molecules for gene therapy.
XX	
PS	Claim 1; SEQ ID NO 824; 205bp; English.
XX	
CC	The present invention relates to novel human secretory or membrane
CC	proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-
CC	ADA54071). The coding sequences are useful in the gene therapy of
CC	diseases caused by abnormalities of the proteins, e.g. cancer,
CC	inflammatory diseases, osteoporosis or neurological disease.
XX	
SQ	Sequence 2100 BP; 461 A; 691 C; 548 G; 400 T; 0 U; 0 Other;
QY	Query Match 38.2%; Score 25.6; DB 10; Length 2100;
Db	Best Local Similarity, 66.1%; Pred. No. 19;
	Matches 37; Conservative 0; Mismatches 19; Indels 0; Gaps 0
3	CAAGCTGGAATTAATTAAGTCTGCCCCACAGTGTGAATTTAGGATTAACAG 58
1414	CCAGCTGGAAGCTGCTCTTCTGCTCCACAGTGTGATCTTTTGAAGAAG 1469
RESULT 12	
ADBI3890	
ID	ADBI3890 standard; DNA; 5054 BP.
XX	
AC	ADBI3890;
XX	
DT	29-JAN-2004 (first entry)
XX	
DE	Human optineurin gene promoter.
XX	
KW	Human; optineurin; ds; promoter; ophthalmological;
KW	single nucleotide polymorphism; SNP; glaucoma;
KW	progressive ocular hypertensive disorder; glaucoma related disorder.
XX	
OS	Homo sapiens.
XX	
Key	Location/Qualifiers
FT	variation
FT	replace(391,G)
FT	/*tag= a
FT	/standard_name= "Single nucleotide polymorphism"
FT	replace(651,G)
FT	/*tag= b
FT	/standard_name= "Single nucleotide polymorphism"
FT	replace(709,G)
FT	/*tag= c
FT	/standard_name= "Single nucleotide polymorphism"
FT	replace(887,A)
FT	/*tag= d
FT	/standard_name= "Single nucleotide polymorphism"
FT	replace(894,T)
FT	/*tag= e
FT	/standard_name= "Single nucleotide polymorphism"
FT	replace(987,C)
FT	/*tag= f
FT	/standard_name= "Single nucleotide polymorphism"
FT	replace(1112,C)
FT	/*tag= g
FT	/standard_name= "Single nucleotide polymorphism"
FT	replace(1505,CC)
FT	/*tag= h
FT	/standard_name= "Single nucleotide polymorphism"
FT	replace(1606,A)
FT	/*tag= i
FT	/standard_name= "Single nucleotide polymorphism"
TT	

FT variation replace(2405,T)  
 FT /<tag= j  
 FT /standard\_name= "Single nucleotide polymorphism"  
 FT replace(2606,G)  
 FT /<tag= k  
 FT /standard\_name= "Single nucleotide polymorphism"  
 FT replace(3313,A)  
 FT /<tag= l  
 FT /standard\_name= "Single nucleotide polymorphism"  
 FT replace(3555,TT)  
 FT /<tag= m  
 FT /standard\_name= "Single nucleotide polymorphism"  
 FT replace(3625,G)  
 FT /<tag= n  
 FT /standard\_name= "Single nucleotide polymorphism"  
 FT replace(3629,C)  
 FT /<tag= o  
 FT /standard\_name= "Single nucleotide polymorphism"  
 FT replace(3882,TT)  
 FT /<tag= p  
 FT /standard\_name= "Single nucleotide polymorphism"  
 FT replace(3988,T)  
 FT /<tag= q  
 FT /standard\_name= "Single nucleotide polymorphism"  
 FT replace(4452,A)  
 FT /<tag= r  
 FT /standard\_name= "Single nucleotide polymorphism"  
 FT /standard\_name= "Single nucleotide polymorphism"  
 FT US2003190617-A1.  
 FT 09-OCT-2003.  
 FT 06-MAR-2002; 2002US-00091281.  
 FT 06-MAR-2002; 2002US-00091281.  
 FT (SIEE/) SI E.  
 FT (RAYM/) RAYMOND V.  
 FT (MORI/) MORISSETTE J.  
 FT Raymond V, Morissette J, SI E;  
 FT MPI; 2003-864168/80.  
 FT New nucleic acid sequences of the optineurin gene are useful to detect  
 FT polymorphisms particularly single nucleotide polymorphisms in the  
 FT optineurin promoter to diagnose, prognose and treat glaucoma and related  
 FT disorders.  
 FT Claim 1, SEQ ID NO 1; 159pp; English.  
 CC The invention relates to an isolated nucleic acid (N1) comprising at  
 CC least 20 but not more than 1500 consecutive nucleotides of the optineurin  
 CC promoter appearing as ADB33890. Also included are the optineurin promoter  
 CC operably linked to a heterologous nucleic acid, a nucleic acid capable of  
 CC detecting a single nucleotide polymorphism (SNP) in the optineurin  
 CC promoter, a host cell comprising the promoter operably linked to a  
 CC heterologous sequence, diagnosing or prognosing glaucoma in a sample  
 CC obtained from a cell or bodily fluid (comprising detecting a polymorphism  
 CC in a promoter region of the optineurin gene, associated with a glaucoma  
 CC phenotype), detecting a SNP sequence variation in a sample containing  
 CC DNA, detecting the presence of an optineurin promoter sequence variation  
 CC in a sample containing DNA, determining the presence or increased  
 CC susceptibility to glaucoma or to a progressive ocular hypertensive  
 CC disorder resulting in loss of visual field in a patient (or the severity  
 CC or progression of glaucoma in a patient, comprising providing  
 CC amplification reaction primers that direct amplification of a selected  
 CC nucleic acid region containing the variation within the optineurin  
 CC promoter and amplifying the DNA) and detecting a polymorphism (comprising  
 CC obtaining a sample containing human genomic DNA, providing a nucleic acid  
 CC capable of detecting a SNP located within an optineurin promoter, and  
 CC detecting the polymorphism). The invention is used to diagnose and  
 CC prognose glaucoma and also to treat glaucoma related disorders. The

CC present sequence is the optineurin promoter.  
 XX SQ Sequence 5054 BP; 1161 A; 1377 C; 1175 G; 1341 T; 0 U; 0 Other;  
 Query Match 38.2%; Score 25.6; DB 10; Length 5054;  
 Best Local Similarity 66.1%; Pred. No. 23;  
 Matches 37; Conservative 0; Mismatches 19; Indels 0; Gaps 0;  
 QY 9 TGGAAATTAAGCTCTCCACCACTTGTAAGTTAGGCTTTAAGACAGTACAG 64  
 DB 3290 TGGGATTACAGGTGAGCCACCGTCCACCTATAGGGTTTAAACAGTAAAG 3345  
 RESULT 13  
 ADE13891 ID ADE13891 standard; DNA; 46951 BP.  
 AC ADE13891;  
 XX 29-JAN-2004 (first entry)  
 DT Human optineurin gene.  
 XX Human; optineurin; ds; gene; ophthalmological;  
 KM single nucleotide polymorphism; SNP; glaucoma;  
 KM progressive ocular hypertensive disorder; glaucoma related disorder.  
 XX Homo sapiens.  
 OS  
 XX Key  
 FH Location/Qualifiers  
 FT variation replace(391,G)  
 FT /<tag= a  
 FT /standard\_name= "Single nucleotide polymorphism"  
 FT replace(691,G)  
 FT /<tag= b  
 FT /standard\_name= "Single nucleotide polymorphism"  
 FT replace(709,G)  
 FT /<tag= c  
 FT /standard\_name= "Single nucleotide polymorphism"  
 FT replace(887,A)  
 FT /<tag= d  
 FT /standard\_name= "Single nucleotide polymorphism"  
 FT replace(894,T)  
 FT /<tag= e  
 FT /standard\_name= "Single nucleotide polymorphism"  
 FT replace(987,C)  
 FT /<tag= f  
 FT /standard\_name= "Single nucleotide polymorphism"  
 FT replace(1112,C)  
 FT /<tag= g  
 FT /standard\_name= "Single nucleotide polymorphism"  
 FT replace(1505,CC)  
 FT /<tag= h  
 FT /standard\_name= "Single nucleotide polymorphism"  
 FT replace(1606,A)  
 FT /<tag= i  
 FT /standard\_name= "Single nucleotide polymorphism"  
 FT replace(2405,T)  
 FT /<tag= j  
 FT /standard\_name= "Single nucleotide polymorphism"  
 FT replace(2606,G)  
 FT /<tag= k  
 FT /standard\_name= "Single nucleotide polymorphism"  
 FT replace(3313,A)  
 FT /<tag= l  
 FT /standard\_name= "Single nucleotide polymorphism"  
 FT replace(3555,TT)  
 FT /<tag= m  
 FT /standard\_name= "Single nucleotide polymorphism"  
 FT replace(3625,G)  
 FT /<tag= n  
 FT /standard\_name= "Single nucleotide polymorphism"  
 FT replace(3629,C)  
 FT variation

```

FT      /tag= 0
FT      /standard_name= "Single nucleotide polymorphism"
FT      replace(3882,TT)
FT      **tag= P
FT      /standard_name= "Single nucleotide polymorphism"
FT      replace(3988,T)
FT      **tag= G
FT      /standard_name= "Single nucleotide polymorphism"
FT      replace(4452,A)
FT      **tag= T
FT      /standard_name= "Single nucleotide polymorphism"
XX      US2003190617-A1.
XX      09-OCT-2003.
XX      06-MAR-2002; 2002US-00091281.
XX      06-MAR-2002; 2002US-00091281.
XX      (SIEE/) SI E.
XX      (RAYM/) RAYMOND V.
XX      (MORI/) MORISSETTE J.
XX      Raymond V, Morissette J, Si E;
XX      WPI; 2003-864168/80.
XX      New nucleic acid sequences of the optineurin gene are useful to detect
XX      polymorphisms particularly single nucleotide polymorphisms in the
XX      optineurin promoter to diagnose, prognose and treat glaucoma and related
XX      disorders.
XX      Disclosure; SEQ ID NO 2; 159pp; English.
XX      The invention relates to an isolated nucleic acid (NI) comprising at
XX      least 20 but not more than 1500 consecutive nucleotides of the optineurin
XX      promoter appearing as ADE13890. Also included are the optineurin promoter
XX      operably linked to a heterologous nucleic acid, a nucleic acid capable of
XX      detecting a single nucleotide polymorphism (SNP) in the optineurin
XX      promoter, a host cell comprising the promoter operably linked to a
XX      heterologous sequence, diagnosing or prognosing glaucoma in a sample
XX      obtained from a cell or bodily fluid (comprising detecting a polymorphism
XX      in a promoter region of the optineurin gene, associated with a glaucoma
XX      phenotype), detecting a SNP sequence variation in a sample containing
XX      DNA, detecting the presence of an optineurin promoter sequence variation
XX      in a sample containing DNA, determining the presence or increased
XX      susceptibility to glaucoma or to a progressive ocular hypertensive
XX      disorder resulting in loss of visual field in a patient (or the severity
XX      or progression of glaucoma in a patient, comprising providing
XX      amplification reaction primers that direct amplification of a selected
XX      nucleic acid region containing the variation within the optineurin
XX      promoter and amplifying the DNA) and detecting a polymorphism (comprising
XX      obtaining a sample containing human genomic DNA, providing a nucleic acid
XX      capable of detecting a SNP located within an optineurin promoter, and
XX      detecting the polymorphism). The invention is used to diagnose and
XX      prognose glaucoma and also to treat glaucoma related disorders. The
XX      present sequence is the optineurin gene.
XX      SQ Sequence 46951 BP; 12703 A; 10108 C; 10051 G; 14089 T; 0 U; 0 Other;
XX      Query Match 38.2%; Score 25.6; DB 10; Length 46951;
XX      Best Local Similarity 66.1%; Pred. No. 39;
XX      Matches 37; Conservative 0; Mismatches 19; Indels 0; Gaps 0;
QY      9 TCGAATAAAGTCTGCCACAGTTGTAAGTTAGGTTTAAGACAGTACAG 64
Db      TGGGATTACAGGTGAGCCACCGTCCCACTATAGGTTTAAACAGTAAAG 3345

```

```

XX      AC ADA30717;
XX      DT 20-NOV-2003 (first entry)
XX      DE DNA encoding Acinetobacter baumannii protein #2004.
XX      KW ds; gene; Acinetobacter baumannii; bacterial disease; antibacterial;
XX      vaccine; plant biocontrol agent.
XX      OS Acinetobacter baumannii.
XX      PN US6562958-B1.
XX      PD 13-MAY-2003.
XX      PF 04-JUN-1999; 99US-00328352.
XX      PR 09-JUN-1998; 98US-0088701P.
XX      PA (GENO-) GENOME THERAPEUTICS CORP.
XX      PI Breton G, Bush D;
XX      PI WPI; 2003-576092/54.
XX      DR P-PSDB; ADA34643.
XX      PT New Acinetobacter baumannii proteins and nucleic acids, useful as reagents
XX      for diagnosing a bacterial disease, as components of antibacterial
XX      vaccines, as targets for antibacterial drugs, or as biocontrol agents for
XX      plants.
XX      Example; SEQ ID NO 2004; 328pp; English.
XX      The invention relates to isolated Acinetobacter baumannii nucleic acids.
XX      CC The A. baumannii nucleic acids and polypeptides are useful as reagents
XX      CC for diagnosing a bacterial disease, as components of antibacterial
XX      CC vaccines, as targets for antibacterial drugs, to detect the presence of
XX      CC A. baumannii and other Acinetobacter species in a sample, in screening
XX      CC compounds for the ability to interfere with the A. baumannii life cycle
XX      CC or to inhibit A. baumannii infection, and as biocontrol agents for
XX      CC plants. The present sequence represents DNA encoding an A. baumannii
XX      CC protein.
XX      SQ Sequence 1284 BP; 368 A; 240 C; 311 G; 365 T; 0 U; 0 Other;
XX      Query Match 37.9%; Score 25.4; DB 9; Length 1284;
XX      Best Local Similarity 61.2%; Pred. No. 20;
XX      Matches 41; Conservative 0; Mismatches 26; Indels 0; Gaps 0;
QY      1 CACAGCTTGGAATTAAGTCTGCCACAGTTGTAAGTTAGGTTTAAGCAGTA 60
Db      CAAAACCTTGACAGCACTTTGGCCTTCAATTGTGATTTTGCAGCACTT 629
QY      61 CAAAGATC 67
Db      628 CACGAC 622

```

```

RESULT 15
ADP49338/C
ID ADP49338 standard; DNA; 80000 BP.
XX      AC ADP49338;
XX      DT 26-AUG-2004 (first entry)
XX      DE Human B-cell chronic lymphocytic leukaemia associated genomic DNA #2.
XX      KW ds; gene; human; B-cell chronic lymphocytic leukaemia; B-CLL; cytostatic;
XX      AMB1; cancer; chromosome 12q21-22.
XX      OS Homo sapiens.

```

```

RESULT 14
ADA30717/C
ID ADA30717 standard; DNA; 1284 BP.

```

```

XX WO2004046376-A2.
PN
XX
XX 03-JUN-2004.
PD
XX
XX 19-NOV-2003; 2003WO-DK000794.
PF
XX
XX 19-NOV-2002; 2002DK-00001792.
PR
XX
XX (HSRI-) HS RIGSHOSPITAL ET.
PA (UYAR-) UNIV ARHUS.
XX
XX Hertz AMB, Leftere H, Kjeme J;
PI
XX WPI; 2004-449754/42.
DR
XX
XX Establishing diagnosis or prognosis of subtype of B-cell chronic
PT lymphocytic leukemia in individual, involves detecting presence or
PT absence of expression product transcribed from human chromosome 12q21-22,
PT in biological sample.
XX
XX Claim 37; Page 73-95; 11pp; English.
PS
XX
XX The present invention relates to a method of establishing diagnosis or
CC prognosis of a subtype of B-cell chronic lymphocytic leukemia (B-CLL) in
CC individual, which involves detecting the presence or absence of an
CC expression product in a biological sample isolated from the individual,
CC where the expression product comprises a nucleotide sequence transcribed
CC from human chromosome 12q21-22. The sequence is designated AMB1. The
CC method is useful for establishing a diagnosis of a subtype of B-CLL in an
CC individual. The sequences can be used for the treatment of cancer, and
CC for the preparation of a medicament for the treatment of cancer. The
CC cancer is B-CLL. The present sequence is a coding sequence from the human
CC chromosome 12q21-22 which is associated with B-CLL.
XX
SQ Sequence 80000 BP; 24867 A; 15522 C; 15290 G; 24321 T; 0 U; 0 Other;

```

Query Match 37.0%; Score 24.8; DB 12; Length 80000;

Best Local Similarity 72.7%; Pred. No. 90;

Matches 32; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 1 CACAACTTGAAATTAAGTCGCCACCACTGTGTAATT 44

Db 72091 CAAAATATTTAATTAATTAATTCGACTGAAACATCTGTAACTT 72048

Search completed: February 8, 2005, 13:57:10  
Job time : 94.8783 secs

**This Page Blank (uspto)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 8, 2005, 13:03:55 ; Search time 18.0955 Seconds  
(without alignments)  
2631.757 Million cell updates/sec

Title: US-10-754-437-8

Perfect score: 67  
Sequence: 1 cacaagctcggaataaag.....gttaagacagctacaagatc 67

Scoring table: IDENTITY\_MNC  
Gapop 10.0 , Gapext 1.0

Searched: 824507 seqs, 355394441 residues

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:  
1: /cgn2\_6/prodata/1/ina/5A.COMB.seq.\*  
2: /cgn2\_6/prodata/1/ina/5B.COMB.seq.\*  
3: /cgn2\_6/prodata/1/ina/6A.COMB.seq.\*  
4: /cgn2\_6/prodata/1/ina/6B.COMB.seq.\*  
5: /cgn2\_6/prodata/1/ina/6C.COMB.seq.\*  
6: /cgn2\_6/prodata/1/ina/backfile1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	26.6	39.7	9262	4 US-09-566-921-74	Sequence 74, Appl
2	25.4	37.9	1284	4 US-09-328-352-2004	Sequence 2004, Ap
3	24.4	36.4	1389	4 US-09-711-164-174	Sequence 174, App
4	24	35.8	2024	4 US-09-643-657-1	Sequence 1, Appli
5	24	35.8	2024	4 US-09-643-657-2	Sequence 2, Appli
6	23.4	34.9	18762	4 US-09-513-999C-18762	Sequence 18762, A
7	23.4	34.9	786431	4 US-09-751-389-3	Sequence 3, Appli
8	23.4	34.9	1830121	4 US-09-557-1884-1	Sequence 1, Appli
9	23.4	34.9	1830121	4 US-09-643-990A-1	Sequence 1, Appli
10	23.4	34.9	1830121	4 US-10-329-960-1	Sequence 1, Appli
11	23.2	34.6	3414	1 US-08-764-100-17	Sequence 17, Appl
12	23.2	34.6	3414	1 US-08-764-100-21	Sequence 21, Appl
13	23.2	34.6	4970	1 US-08-764-100-14	Sequence 14, Appl
14	23.2	34.6	4970	1 US-08-764-100-20	Sequence 20, Appl
15	23	34.6	268	4 US-09-513-999C-27534	Sequence 27534, A
16	23	34.3	1435	5 PCT-US95-05922A-1	Sequence 1, Appli
17	23	34.3	1601	4 US-08-023-655-497	Sequence 497, App
18	23	34.3	2580	4 US-08-511-485-7	Sequence 7, Appli
19	23	34.3	2580	4 US-09-201-936-7	Sequence 7, Appli
20	23	34.3	2580	4 US-09-011-356-7	Sequence 7, Appli
21	23	34.3	2580	4 US-09-672-717-222	Sequence 222, App
22	23	34.3	2580	4 US-09-201-932-7	Sequence 7, Appli
23	23	34.3	2588	3 US-08-569-749-1	Sequence 1, Appli
24	23	34.3	2589	5 PCT-US96-12860-1	Sequence 1, Appli
25	23	34.3	3171	2 US-08-868-786-5	Sequence 5, Appli
26	23	34.3	3532	2 US-09-205-204-1	Sequence 1, Appli
27	23	34.3	3732	3 US-09-212-971-7	Sequence 7, Appli

## ALIGNMENTS

28	23	34.3	3732	3 US-08-800-929A-7	Sequence 7, Appli
29	23	34.3	3732	3 US-09-617-053A-7	Sequence 7, Appli
30	22.8	34.0	336	4 US-09-640-173-35	Sequence 35, Appl
31	22.8	34.0	336	4 US-09-713-550-35	Sequence 35, Appl
32	22.8	34.0	336	4 US-09-825-294-35	Sequence 35, Appl
33	22.8	34.0	336	4 US-09-970-966-35	Sequence 35, Appl
34	22.8	34.0	892	4 US-09-270-767-2623	Sequence 2623, Ap
35	22.8	34.0	892	4 US-09-270-767-17905	Sequence 17905, A
36	22.8	34.0	1055	4 US-09-270-767-9711	Sequence 9711, Ap
37	22.8	34.0	1055	4 US-09-270-767-24993	Sequence 24993, A
38	22.8	34.0	1249	4 US-09-220-132-4	Sequence 4, Appli
39	22.6	33.7	378	4 US-09-288-143-39	Sequence 39, Appl
40	22.4	33.4	202	4 US-09-513-999C-33953	Sequence 33953, A
41	22.4	33.4	843	4 US-09-248-796A-3158	Sequence 3158, Ap
42	22.4	33.4	957	4 US-09-134-000C-195	Sequence 195, App
43	22.4	33.4	1861	4 US-09-129-668-7	Sequence 7, Appli
44	22.4	33.4	2065	4 US-09-129-668-5	Sequence 5, Appli
45	22.4	33.4	2084	4 US-09-222-938A-9	Sequence 9, Appli

  

RESULT 1	US-09-566-921-74/C
Sequence 74, Application US/09566921	
Patent No. 6682888	
GENERAL INFORMATION:	
APPLICANT: Loring, Jeanne F.	
APPLICANT: Tingley, Debora W.	
APPLICANT: Edwards, Carla M.	
TITLE OF INVENTION: GENES EXPRESSED IN ALZHEIMER'S DISEASE	
FILE REFERENCE: PA-0024 US	
CURRENT APPLICATION NUMBER: US/09/566,921	
CURRENT FILING DATE: 2000-05-05	
NUMBER OF SEQ ID NOS: 138	
SOFTWARE: PERL Program	
SEQ ID NO 74	
LENGTH: 9262	
TYPE: DNA	
ORGANISM: Homo sapiens	
FEATURE:	
NAME/KEY: misc feature	
OTHER INFORMATION: Incycle ID No. 6682888 246655.92	
NAME/KEY: unsure	
LOCATION: 1941-1942, 1952-1953, 1956, 1958, 1963, 6600-6601	
OTHER INFORMATION: a, t, c, g, or other	
US-09-566-921-74	
Query Match	39.7%; Score 26.6; DB 4; Length 9262;
Best Local Similarity	63.1%; Pred. No. 1.3;
Matches 41; Conservative	0; Mismatches 24; Indels 0; Gaps 0;
DB	3 CAGCTTGAAATTAAGTGTGCCACAGTGTGACTTGGGTTAAGACAGTACA 62
DB	5902 CATTAAATGATTAATTAATGTTGCCCCAGTGTATCTTAAGATTAAGACTATCA 5843
QY	63 AGATC 67
DB	5842 TGATC 5838
RESULT 2	US-09-328-352-2004/C
Sequence 2004, Application US/09328352	
Patent No. 6562958	
GENERAL INFORMATION:	
APPLICANT: Gary L. Breton et al.	
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER	
FILE REFERENCE: BAUMANNI FOR DIAGNOSTICS AND THERAPEUTICS	
FILE REFERENCE: GTC99-03PA	
CURRENT APPLICATION NUMBER: US/09/328,352	
CURRENT FILING DATE: 1999-06-04	

NUMBER OF SEQ ID NOS: 8252  
SEQ ID NO 2004  
LENGTH: 1284  
TYPE: DNA  
ORGANISM: Acinetobacter baumannii  
US-09-328-352-2004

Query Match 37.9%; Score 25.4; DB 4; Length 1284;  
Best Local Similarity 61.2%; Pred. No. 2.1;  
Matches 41; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 1 CACAAGCTTGGAAATTAAGTCTGCCACCACTTGTAACTTTAGGTTTAAAGACAGTA 60  
DB 688 CAAAGCTTGAAGACGAACTTTGCGCTTCAATTGTAGTATTTTTCAGCAACTT 629

QY 61 CAGATC 67  
DB 628 CACGACC 622

## RESULT 3

US-09-711-164-174/C  
Sequence 174, Application US/09711164  
Patent No. 6589738  
GENERAL INFORMATION:  
APPLICANT: Forsyth, R. Allyn  
APPLICANT: Ohlsen, Kari  
APPLICANT: Zvekind, Judith  
TITLE OF INVENTION: GENES ESSENTIAL FOR MICROBIAL PROLIFERATION AND ANTISENSE THERAPY  
FILE REFERENCE: ELITRA.008A  
CURRENT APPLICATION NUMBER: US/09/711.164  
PRIOR FILING DATE: 2000-11-09  
PRIOR APPLICATION NUMBER: US 60/164415  
PRIOR FILING DATE: 1999-11-9  
NUMBER OF SEQ ID NOS: 469  
SOFTWARE: FaSeq for Windows Version 4.0  
SEQ ID NO 174  
LENGTH: 1389  
TYPE: DNA  
ORGANISM: Escherichia coli  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)...(1389)  
US-09-711-164-174

Query Match 36.4%; Score 24.4; DB 4; Length 1389;  
Best Local Similarity 60.6%; Pred. No. 5.3;  
Matches 40; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 1 CACAAGCTTGGAAATTAAGTCTGCCACCACTTGTAACTTTAGGTTTAAAGACAGTA 60  
DB 1230 CATCAGCCCGAAGCGAGTACGCCACAGCGGTTGTATGTCGCGTTTACGACACCA 1171

QY 61 CAGAT 66  
DB 1170 GATGAT 1165

## RESULT 4

US-09-643-657-1  
Sequence 1, Application US/09643657  
Patent No. 6642024  
GENERAL INFORMATION:  
APPLICANT: Diane Pennica  
TITLE OF INVENTION: GUANYLATE-BINDING PROTEIN  
NUMBER OF SEQUENCES: 43  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 1 DNA Way  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080

## COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Winpatin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/643,657  
FILING DATE: 17-Aug-2000  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/015,089A  
FILING DATE: 29-Jan-1998  
ATTORNEY/AGENT INFORMATION:  
NAME: Hasak, Janet E.  
REGISTRATION NUMBER: 28,616  
REFERENCE/DOCKET NUMBER: P1056  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650/225-1896  
TELEFAX: 650/952-9881  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2024 base pairs  
TYPE: Nucleic Acid  
STRANDEDNESS: Single  
TOPOLOGY: Linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-09-643-657-1

Query Match 35.8%; Score 24; DB 4; Length 2024;  
Best Local Similarity 64.3%; Pred. No. 8.5;  
Matches 36; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 4 AAGCTTGGAAATTAAGTCTGCCACCACTTGTAACTTTAGGTTTAAAGACAGT 59  
DB 1937 AAGTCTGGTATGAATCCAAAGCTCCACTTGTAGTATATGTGTTAGGCAAGT 1992

## RESULT 5

US-09-643-657-2/C  
Sequence 2, Application US/09643657  
Patent No. 6642024  
GENERAL INFORMATION:  
APPLICANT: Diane Pennica  
TITLE OF INVENTION: GUANYLATE-BINDING PROTEIN  
NUMBER OF SEQUENCES: 43  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 1 DNA Way  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Winpatin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/643,657  
FILING DATE: 17-Aug-2000  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/015,089A  
FILING DATE: 29-Jan-1998  
ATTORNEY/AGENT INFORMATION:  
NAME: Hasak, Janet E.  
REGISTRATION NUMBER: 28,616  
REFERENCE/DOCKET NUMBER: P1056  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650/225-1896  
TELEFAX: 650/952-9881  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:

LENGTH: 2024 base pairs  
TYPE: Nucleic Acid  
STRANDEDNESS: Single  
TOPOLOGY: Linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
US-09-643-657-2

Query Match 35.8%; Score 24; DB 4; Length 2024;  
Best Local Similarity 64.3%; Pred. No. 8.5;  
Matches 36; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 4 AACCTTGGAATTAAGCTGCCCCCAGTGGTAAGGTTTAAGACAGT 59  
DB 88 AAGCTCGGTAATGATCCAGCTCCACTCTTAAGTAATAGTTAGCAAGT 33

RESULT 6  
US-09-513-999C-18762/C  
Sequence 18762, Application US/09513999C  
Patent No. 6783961

GENERAL INFORMATION:  
APPLICANT: Dumas Milne Edwards, J.B.  
APPLICANT: Duclert, A.  
APPLICANT: Giordano, J.Y.  
TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.  
Patent No. 6783961

FILE REFERENCE: 59.US2.REG  
CURRENT APPLICATION NUMBER: US/09/513,999C  
CURRENT FILING DATE: 2000-02-24  
PRIOR APPLICATION NUMBER: US 60/122,487  
PRIOR FILING DATE: 1999-02-26  
NUMBER OF SEQ ID NOS: 36681  
SOFTWARE: Patent.pm  
SEQ ID NO 18762

LENGTH: 143  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 116  
OTHER INFORMATION: k=g or t  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 117  
OTHER INFORMATION: r=a or g  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 125  
OTHER INFORMATION: m=a or c  
US-09-513-999C-18762

Query Match 34.9%; Score 23.4; DB 4; Length 143;  
Best Local Similarity 60.0%; Pred. No. 6.6;  
Matches 39; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 3 CAAGCTTGGAATTAAGCTGCCCCCAGTGGTAAGGTTTAAGACAGT 62  
DB 91 CAATCTTCTTAATATGACCAATTCAGTTTAACTACTGTAATATCAATT 32

QY 63 AGATC 67  
DB 31 AATTC 27

RESULT 7  
US-09-751-389-3/C  
Sequence 3, Application US/09751389  
Patent No. 6630334  
GENERAL INFORMATION:  
APPLICANT: GUGELER, Karl et al  
TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC  
ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES  
TITLE OF INVENTION: THEREOF

FILE REFERENCE: CL001067  
CURRENT APPLICATION NUMBER: US/09/751,389  
CURRENT FILING DATE: 2001-01-02  
NUMBER OF SEQ ID NOS: 8  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 3  
LENGTH: 786431  
TYPE: DNA  
ORGANISM: Human  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(786431)  
OTHER INFORMATION: n = A,T,C or G  
US-09-751-389-3

Query Match 34.9%; Score 23.4; DB 4; Length 786431;  
Best Local Similarity 63.2%; Pred. No. 80;  
Matches 36; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 8 TTGGAATTAAGCTGCCCCCAGTGGTAAGGTTTAAGACAGTCAAG 64  
DB 62963 TTGGCAATATTAAGCTTCCGAAAGGACAGAAATTTCTGTTAAGAGCAATAG 62907

RESULT 8  
US-09-557-884-1/C  
Sequence 1, Application US/09557884  
Patent No. 6506581

GENERAL INFORMATION:  
APPLICANT: Fleischmann et al.  
TITLE OF INVENTION: The Nucleotide sequence of  
the Haemophilus influenzae Rd Genome, Fragments  
Thereof, and Uses Thereof  
NUMBER OF SEQUENCES: 1  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: MD  
COUNTRY: USA  
ZIP: 20850

COMPUTER READABLE FORM:  
MEDIUM TYPE: 3 1/2 inch diskette  
COMPUTER: Dell Pentium  
OPERATING SYSTEM: MS DOS v6.22  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/557,884  
FILING DATE: 25-Apr-2000  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/476,102  
FILING DATE: JUN-5-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Michelle S. Marks  
REGISTRATION NUMBER: 41,971  
REFERENCE/DOCKET NUMBER: PB186P3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 301-309-8504  
TELEFAX: 301-309-8439

INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1830121 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-09-557-884-1

Query Match 34.9%; Score 23.4; DB 4; Length 1830121;  
Best Local Similarity 60.0%; Pred. No. 93;  
Matches 39; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 1 CACAGCTTGGAAATAAAGTGTGCCCCACAGTTGGTAACCTTAGGGTTAAGACAGTA 60  
Db 1518906 CATACGTTTACCAATAAATTGCCCCACAGCTTGGTTGTTGTGACACACCGTG 1518847

QY 61 CAAGA 65  
Db 1518846 TAATA 1518842

## RESULT 9

US-09-643-990A-1/c  
; Sequence 1, Application US/09643990A  
; Patent No. 6528289  
; GENERAL INFORMATION:  
; APPLICANT: Robert D. Fleischmann  
; Mark D. Adams  
; Owen White  
; Hamilton O. Smith  
; J. Craig Venter  
; TITLE OF INVENTION: The Nucleotide sequence of  
the Haemophilus influenzae Rd Genome, Fragments  
Thereof, and Uses Thereof

NUMBER OF SEQUENCES: 1  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville,  
STATE: MD  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3 1/2 inch diskette  
COMPUTER: Dell Pentium  
OPERATING SYSTEM: MS DOS V6.22  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/643,990A  
FILING DATE: 23-Aug-2000  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/487,429  
FILING DATE: 1995-06-07  
APPLICATION NUMBER: 08/426,787  
FILING DATE: 1995-04-21  
ATTORNEY/AGENT INFORMATION:  
NAME: Kenley K. Hoover  
REGISTRATION NUMBER: 40,302  
REFERENCE/DOCKET NUMBER: PB186P1C1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 301-610-5790  
TELEFAX: 310-309-8439  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1830121 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-09-643-990A-1

Query Match 34.9%; Score 23.4; DB 4; Length 1830121;  
Best Local Similarity 60.0%; Pred. No. 93;  
Matches 39; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 1 CACAGCTTGGAAATAAAGTGTGCCCCACAGTTGGTAACCTTAGGGTTAAGACAGTA 60  
Db 1518906 CATACGTTTACCAATAAATTGCCCCACAGCTTGGTTGTTGTGACACACCGTG 1518847

QY 61 CAAGA 65  
Db 1518846 TAATA 1518842

## RESULT 10

US-10-329-960-1/c  
; Sequence 1, Application US/10329960  
; Patent No. 6742927  
; GENERAL INFORMATION:  
; APPLICANT: Fleischmann et al.  
; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus influenzae Rd Genome, Fra  
ment No. 6742927  
; TITLE OF INVENTION: Thereof, and Uses Thereof  
; FILE REFERENCE: PB186P1  
; CURRENT APPLICATION NUMBER: US/10/329,960  
; CURRENT FILING DATE: 2003-01-02  
; PRIOR APPLICATION NUMBER: US 09/643,990  
; PRIOR FILING DATE: 2000-08-23  
; PRIOR APPLICATION NUMBER: US 08/487,429  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: US 08/426,787  
; PRIOR FILING DATE: 1995-04-21  
; NUMBER OF SEQ ID NOS: 1  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 1

LENGTH: 1830121

TYPE: DNA  
ORGANISM: Haemophilus influenzae

FEATURE:  
NAME/KEY: misc feature  
LOCATION: (4747)..(4747)  
OTHER INFORMATION: n equals a, t, g or c

FEATURE:  
NAME/KEY: misc feature  
LOCATION: (9921)..(9921)  
OTHER INFORMATION: n equals a, t, g or c

FEATURE:  
NAME/KEY: misc feature  
LOCATION: (10150)..(10150)  
OTHER INFORMATION: n equals a, t, g or c

FEATURE:  
NAME/KEY: misc feature  
LOCATION: (29298)..(29298)  
OTHER INFORMATION: n equals a, t, g or c

FEATURE:  
NAME/KEY: misc feature  
LOCATION: (36543)..(36543)  
OTHER INFORMATION: n equals a, t, g or c

FEATURE:  
NAME/KEY: misc feature  
LOCATION: (36551)..(36551)  
OTHER INFORMATION: n equals a, t, g or c

FEATURE:  
NAME/KEY: misc feature  
LOCATION: (36636)..(36636)  
OTHER INFORMATION: n equals a, t, g or c

FEATURE:  
NAME/KEY: misc feature  
LOCATION: (40808)..(40810)  
OTHER INFORMATION: n equals a, t, g or c

FEATURE:  
NAME/KEY: misc feature  
LOCATION: (44416)..(44416)  
OTHER INFORMATION: n equals a, t, g or c

FEATURE:  
NAME/KEY: misc feature  
LOCATION: (44905)..(44905)  
OTHER INFORMATION: n equals a, t, g or c

FEATURE:  
NAME/KEY: misc feature  
LOCATION: (44975)..(44975)  
OTHER INFORMATION: n equals a, t, g or c

FEATURE:  
NAME/KEY: misc feature  
LOCATION: (45553)..(45553)  
OTHER INFORMATION: n equals a, t, g or c

FEATURE:  
NAME/KEY: misc feature  
LOCATION: (45593)..(45593)  
OTHER INFORMATION: n equals a, t, g or c

FEATURE:  
NAME/KEY: misc feature  
LOCATION: (45593)..(45593)  
OTHER INFORMATION: n equals a, t, g or c

FEATURE:  
NAME/KEY: misc feature  
LOCATION: (45593)..(45593)  
OTHER INFORMATION: n equals a, t, g or c

FEATURE:  
NAME/KEY: misc feature  
LOCATION: (45593)..(45593)  
OTHER INFORMATION: n equals a, t, g or c

FEATURE:  
NAME/KEY: misc feature  
LOCATION: (45593)..(45593)  
OTHER INFORMATION: n equals a, t, g or c

FEATURE:  
NAME/KEY: misc feature  
LOCATION: (45593)..(45593)  
OTHER INFORMATION: n equals a, t, g or c

FEATURE:  
NAME/KEY: misc feature  
LOCATION: (45593)..(45593)  
OTHER INFORMATION: n equals a, t, g or c

FEATURE:  
NAME/KEY: misc feature  
LOCATION: (45593)..(45593)  
OTHER INFORMATION: n equals a, t, g or c

FEATURE:  
NAME/KEY: misc feature  
LOCATION: (45593)..(45593)  
OTHER INFORMATION: n equals a, t, g or c

FEATURE:  
NAME/KEY: misc feature  
LOCATION: (45593)..(45593)  
OTHER INFORMATION: n equals a, t, g or c

```
? NAME/KEY: misc_feature
? LOCATION: (45732)..(45732)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (47036)..(47036)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (51334)..(51334)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (51602)..(51602)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (51786)..(51786)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (51805)..(51805)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (55369)..(55369)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (65309)..(65309)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (65313)..(65313)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (80024)..(80024)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (100091)..(100091)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (102696)..(102696)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (105121)..(105121)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (107248)..(107248)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (117136)..(117136)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (119244)..(119244)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (120038)..(120038)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
```

```
? LOCATION: (121344)..(121344)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (122167)..(122167)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (122336)..(122336)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (131340)..(131340)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (131360)..(131360)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (139910)..(139910)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (140398)..(140398)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (142750)..(142750)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (145058)..(145058)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (145171)..(145171)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (145942)..(145942)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (147197)..(147197)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (150841)..(150841)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (152500)..(152500)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
```

```
? Query Match 34.9%; Score 23.4; DB 4; Length 1830121;
? Best Local Similarity 60.0%; Pred. No. 93;
? Matches 39; Conservative 0; Mismatches 26; Indels 0; Gaps 0;
?
? QY 1 CACAGCTTGAATTAAGTCTGCCACCACTTGTACTTATAGAGCACTA 60
? DB 1518906 CATACGTTAGCCAAATTAATTGCTTAACCTTGTTGTGTGCACACCGTG 1518847
? QY 61 CAAGA 65
? DB 1518846 TAATA 1518842
```

```
? RESULT 11
? US-08-764-100-17/c
? Sequence 17, Application US/08764100
```

Patent No. 5773700  
GENERAL INFORMATION:  
APPLICANT: van Grinsven J., Martinus Q.  
APPLICANT: De Haan, Petrus T.  
APPLICANT: Gielen L., Johannes J.  
APPLICANT: Peters, Dirk  
APPLICANT: Goldbach, Robert W.  
TITLE OF INVENTION: Improvements in or Relating to Organic  
TITLE OF INVENTION: Compounds  
NUMBER OF SEQUENCES: 27  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Sandoz Agro, Inc  
STREET: 975 California Avenue  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/764,100  
FILING DATE: 06-DEC-1996  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/214,064  
FILING DATE:  
APPLICATION NUMBER: US 08/032,235  
FILING DATE: 17-MAR-1993  
APPLICATION NUMBER: GB 9206016.9  
FILING DATE: 19-MAR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 57737001s, Allen E.  
REGISTRATION NUMBER: 34,490  
REFERENCE/DOCKET NUMBER: 137-1061  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 354-3592  
TELEFAX: (415) 857-1125  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3414 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-764-100-17

Query Match 34.6%; Score 23.2; DB 1; Length 3414;  
Best Local Similarity 70.5%; Pred. No. 20;  
Matches 31; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 11 GAATTAAGTCTGCCCCCAGTGTGTAACCTTTAGGGTTAG 54  
DB 304 GCATATATATCTGCACATCATGTTGGAGCTTTGGAGCATATG 261

RESULT 12  
US-08-764-100-21  
Sequence 21, Application US/08764100  
Patent No. 5773700  
GENERAL INFORMATION:  
APPLICANT: van Grinsven J., Martinus Q.  
APPLICANT: De Haan, Petrus T.  
APPLICANT: Gielen L., Johannes J.  
APPLICANT: Peters, Dirk  
APPLICANT: Goldbach, Robert W.  
TITLE OF INVENTION: Improvements in or Relating to Organic  
TITLE OF INVENTION: Compounds  
NUMBER OF SEQUENCES: 27  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Sandoz Agro, Inc  
STREET: 975 California Avenue

CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/764,100  
FILING DATE: 06-DEC-1996  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/214,064  
FILING DATE:  
APPLICATION NUMBER: US 08/032,235  
FILING DATE: 17-MAR-1993  
APPLICATION NUMBER: GB 9206016.9  
FILING DATE: 19-MAR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 57737001s, Allen E.  
REGISTRATION NUMBER: 34,490  
REFERENCE/DOCKET NUMBER: 137-1061  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 354-3592  
TELEFAX: (415) 857-1125  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3414 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-764-100-21

Query Match 34.6%; Score 23.2; DB 1; Length 3414;  
Best Local Similarity 70.5%; Pred. No. 20;  
Matches 31; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 11 GAATTAAGTCTGCCCCCAGTGTGTAACCTTTAGGGTTAG 54  
DB 311 GCATATATATCTGCACATCATGTTGGAGCTTTGGAGCATATG 3154

RESULT 13  
US-08-764-100-14/C  
Sequence 14, Application US/08764100  
Patent No. 5773700  
GENERAL INFORMATION:  
APPLICANT: van Grinsven J., Martinus Q.  
APPLICANT: De Haan, Petrus T.  
APPLICANT: Gielen L., Johannes J.  
APPLICANT: Peters, Dirk  
APPLICANT: Goldbach, Robert W.  
TITLE OF INVENTION: Improvements in or Relating to Organic  
TITLE OF INVENTION: Compounds  
NUMBER OF SEQUENCES: 27  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Sandoz Agro, Inc  
STREET: 975 California Avenue  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/764,100  
FILING DATE: 06-DEC-1996  
CLASSIFICATION: 800

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/214,064  
FILING DATE: US 08/032,235  
APPLICATION NUMBER: US 08/032,235  
FILING DATE: 17-MAR-1993  
APPLICATION NUMBER: GB 9206016.9  
FILING DATE: 19-MAR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 577370018, Allen E.  
REGISTRATION NUMBER: 34,490  
REFERENCE/DOCKET NUMBER: 137-1061  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 354-3592  
TELEFAX: (415) 857-1125  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4970 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-764-100-14

Query Match 34.6%; Score 23.2; DB 1; Length 4970;  
Best Local Similarity 70.5%; Pred. No. 23;  
Matches 31; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 11 GAATAAAGTCGCCCCCAGCTGTGTAACCTTAGGCTTAAG 54  
DB 1774 GCAATATTAATTCGACATCAGTTGGAGCTTTGGAGCTAAG 1731

RESULT 14  
US-08-764-100-20  
Sequence 20, Application US/08764100  
Patent No. 5773700  
GENERAL INFORMATION:  
APPLICANT: van Griensven J., Martinus Q.  
APPLICANT: De Haan, Petrus T.  
APPLICANT: Gielens L., Johannes J.  
APPLICANT: Peters, Dirk  
APPLICANT: Goldbach, Robert W.  
TITLE OF INVENTION: Improvements in or Relating to Organic  
TITLE OF INVENTION: Compounds  
NUMBER OF SEQUENCES: 27  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sandoz Agro, Inc  
STREET: 975 California Avenue  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/764,100  
FILING DATE: 06-DEC-1996  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/214,064  
FILING DATE: US 08/032,235  
FILING DATE: 17-MAR-1993  
APPLICATION NUMBER: GB 9206016.9  
FILING DATE: 19-MAR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 577370018, Allen E.  
REGISTRATION NUMBER: 34,490  
REFERENCE/DOCKET NUMBER: 137-1061  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 354-3592

TELEFAX: (415) 857-1125  
INFORMATION FOR SEQ ID NO: 20:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4970 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-764-100-20

Query Match 34.6%; Score 23.2; DB 1; Length 4970;  
Best Local Similarity 70.5%; Pred. No. 23;  
Matches 31; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 11 GAATAAAGTCGCCCCCAGCTGTGTAACCTTAGGCTTAAG 54  
DB 3197 GCAATATTAATTCGACATCAGTTGGAGCTTTGGAGCTAAG 3240

RESULT 15  
US-09-513-999C-27534/C  
Sequence 27534, Application US/09513999C  
Patent No. 6783961  
GENERAL INFORMATION:  
APPLICANT: Dumas Milne Edwards, J.B.  
APPLICANT: Duclert, A.Y.  
APPLICANT: Giordano, J.Y.  
TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.  
FILE REFERENCE: 59, US2, REG  
CURRENT APPLICATION NUMBER: US/09/513,999C  
CURRENT FILING DATE: 2000-02-24  
PRIOR APPLICATION NUMBER: US 60/122,487  
PRIOR FILING DATE: 1999-02-26  
NUMBER OF SEQ ID NOS: 36681  
SOFTWARE: Patent.pm  
SEQ ID NO 27534  
LENGTH: 268  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 250  
OTHER INFORMATION: n=a, g, c or t  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 251  
OTHER INFORMATION: d=a or g or t  
US-09-513-999C-27534

Query Match 34.3%; Score 23; DB 4; Length 268;  
Best Local Similarity 63.6%; Pred. No. 11;  
Matches 35; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 11 GAATAAAGTCGCCCCCAGCTGTGTAACCTTAGGCTTAAGAGTACAGAG 65  
DB 113 GAATAAAGTATAGGCCCCCATTTATGGGAAATTTGTATTAAACATTTCAATA 59

Search completed: February 8, 2005, 22:26:12  
Job time : 26.2066 secs

**This Page Blank (uspto)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 8, 2005, 16:37:21 ; Search time 90.2823 Seconds  
(without alignments)  
4264.119 Million cell updates/sec

Title: US-10-754-437-8

Perfect score: 67

Sequence: 1 cacaagcttggaataaag.....gttaagacagtcacagatc 67

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4300275 seqs, 2872944193 residues

Total number of hits satisfying chosen parameters: 8600550

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications NA:\*

1: /cgn2\_6/prodata/2/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/prodata/2/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/prodata/2/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/prodata/2/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/prodata/2/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/prodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/prodata/2/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/prodata/2/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/prodata/2/pubpna/US09A\_PUBCOMB.seq.\*  
10: /cgn2\_6/prodata/2/pubpna/US09B\_PUBCOMB.seq.\*  
11: /cgn2\_6/prodata/2/pubpna/US09C\_PUBCOMB.seq.\*  
12: /cgn2\_6/prodata/2/pubpna/US09\_NEW\_PUB.seq.\*  
13: /cgn2\_6/prodata/2/pubpna/US10\_PUBCOMB.seq.\*  
14: /cgn2\_6/prodata/2/pubpna/US10\_PUBCOMB.seq.\*  
15: /cgn2\_6/prodata/2/pubpna/US10C\_PUBCOMB.seq.\*  
16: /cgn2\_6/prodata/2/pubpna/US10E\_PUBCOMB.seq.\*  
17: /cgn2\_6/prodata/2/pubpna/US10F\_PUBCOMB.seq.\*  
18: /cgn2\_6/prodata/2/pubpna/US10\_NEW\_PUB.seq.\*  
19: /cgn2\_6/prodata/2/pubpna/US10\_NEW\_PUB.seq.\*  
20: /cgn2\_6/prodata/2/pubpna/US60\_NEW\_PUB.seq.\*  
21: /cgn2\_6/prodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	35	52.2	600	US-10-027-632-287050	Sequence 287050,
C 2	35	52.2	600	US-10-027-632-287050	Sequence 287050,
C 3	33.8	50.4	600	US-10-027-632-287049	Sequence 287049,
C 4	33.8	50.4	600	US-10-027-632-287049	Sequence 287049,
C 5	33.8	50.4	259202	US-10-723-860-1311	Sequence 1311, Ap
C 6	26.6	39.7	216	US-10-066-543-2458	Sequence 2458, Ap
C 7	26.6	39.7	559	US-10-060-036-3636	Sequence 3636, Ap
C 8	26.6	39.7	614	US-09-969-034-4304	Sequence 4304, Ap
C 9	26.6	39.7	2516	US-10-723-860-5715	Sequence 5715, Ap
C 10	26.6	39.7	2516	US-10-723-860-5715	Sequence 5715, Ap
C 11	26.6	39.7	9277	US-10-084-817-95	Sequence 95, Appl
C 12	26.4	39.4	617	US-10-027-632-282653	Sequence 282653,

13	26.4	39.4	617	13	US-10-027-632-282654	Sequence 282654,
14	26.4	39.4	617	15	US-10-027-632-282653	Sequence 282653,
15	26.4	39.4	617	15	US-10-027-632-282654	Sequence 282654,
16	26.4	39.4	89856	17	US-10-322-281-79	Sequence 79, Appl
17	26.2	39.1	47804	15	US-10-017-161-717	Sequence 717, Appl
18	26.2	39.1	47804	15	US-10-292-799-629	Sequence 629, Appl
19	26	38.8	1467	16	US-10-425-114-10667	Sequence 10667, A
20	26	38.8	3170	16	US-10-424-599-105269	Sequence 105269,
21	26	38.8	392112	18	US-10-812-232-3	Sequence 3, Appl1
22	25.6	38.2	455	10	US-09-918-995-295	Sequence 295, Appl
23	25.6	38.2	2100	15	US-10-094-749-824	Sequence 824, Appl
24	25.6	38.2	5054	15	US-10-091-281-1	Sequence 1, Appl1
25	25.6	38.2	44728	17	US-10-367-094-23	Sequence 23, Appl
26	25.6	38.2	46951	15	US-10-091-281-2	Sequence 2, Appl1
27	25.4	37.9	30752	13	US-10-087-192-745	Sequence 745, Appl
28	25	37.3	413	16	US-10-424-599-76447	Sequence 76447, A
29	24.8	37.0	2207	17	US-10-437-963-10374	Sequence 10374, A
30	24.6	36.7	460	16	US-10-424-599-83824	Sequence 83824, A
31	24.6	36.7	470	9	US-09-864-761-1997	Sequence 1997, Ap
32	24.6	36.7	515	15	US-10-029-386-11997	Sequence 11997, A
33	24.6	36.7	879	16	US-10-296-115-595	Sequence 595, Appl
34	24.6	36.7	1147	13	US-10-027-632-118119	Sequence 118119,
35	24.6	36.7	1147	15	US-10-027-632-118119	Sequence 118119,
36	24.6	36.7	49031	17	US-10-322-281-523	Sequence 523, Appl
37	24.6	36.7	65359	9	US-09-804-472-3	Sequence 3, Appl1
38	24.4	36.4	487	9	US-09-292-758-84	Sequence 84, Appl
39	24.4	36.4	1389	15	US-10-287-274-114	Sequence 174, Appl
40	24.4	36.4	1389	15	US-10-369-493-47263	Sequence 47263, A
41	24.4	36.4	1389	16	US-10-282-122A-6469	Sequence 6469, Ap
42	24.4	36.4	27893	15	US-10-017-161-757	Sequence 757, Appl
43	24.2	36.1	376	16	US-10-424-599-72155	Sequence 72155, A
44	24.2	36.1	383	10	US-09-918-995-18275	Sequence 18275, A
45	24.2	36.1	1242	18	US-10-723-860-6348	Sequence 6348, Ap

#### ALIGNMENTS

RESULT 1  
US-10-027-632-287050/c  
; Sequence 287050, Application US/10027632  
; Publication No. US20020198371A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 287050  
; LENGTH: 600  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-287050

Query Match 52.2%; Score 35; DB 13; Length 600;  
Best Local Similarity 79.1%; Pred. No. 0.0044;

Matches 53; Conservative 1; Mismatches 11; Indels 2; Gaps 1;

```

Oy      1  CACAAAGCTTGGAAATATAAAGCTGCCCCACCAAGTTGGTAACCTTTAGGGCTTTAAACACAGTA  60
        |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      163  CACAAAGCTTGGACATAAAATCTGCYC- -ATAGTTGGTGAATTAGGGTTTAAACAGTA  108

```

QY	61	CAAGATC	67
Db	105	TGAGATC	99

## RESULT 2

US-10-027-632-287050/c  
; Sequence 287050, Application US/10027632  
; Publication No. US20030204075A9  
GENERAL INFORMATION:

```

APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
POLYMORPHISMS IN THE HUMAN GENOME
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIORITY APPLICATION NUMBER: US 60/218,006
PRIORITY FILING DATE: 2000-07-12
PRIORITY APPLICATION NUMBER: US 60/198,676
PRIORITY FILING DATE: 2000-04-20
PRIORITY APPLICATION NUMBER: US 60/193,483
PRIORITY FILING DATE: 2000-03-29
PRIORITY APPLICATION NUMBER: US 60/185,218
PRIORITY FILING DATE: 2000-02-24
PRIORITY APPLICATION NUMBER: US 60/167,363
PRIORITY FILING DATE: 1999-11-23
PRIORITY APPLICATION NUMBER: US 60/156,358
PRIORITY FILING DATE: 1999-09-28
PRIORITY APPLICATION NUMBER: US 60/146,002
PRIORITY FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 287050
LENGTH: 600
TYPE: DNA
ORGANISM: Human
US-10-027-632-287050

```

Query Match	52.2%;	Score 35;	DB 15;	Length 600;
Best Local Similarity	79.1%;	Pred. No. 0.0044;		
Matches 53; Conservative	1;	Mismatches 11;	Indels 2;	Gaps 1

OY 1 CACAAGCTTGGAAATTAAGAAGTCGCCCCACCAAGTTGTACTTTAGGGTTTAAGACAGTA 60  
||| ||| ||| ||| : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 163 CACAAACTTGACATATAAAAATCTGC - ATAGTTGGTAATTAGCGTTTAAACAGTA 106

QY	61	CAAGATC	67
Db	105	TGAGATC	99

### RESULT 3

US-10-027-632-287049/C  
; Sequence 287049, Application US/10027632  
; Publication No. US20020198371A1  
GENERAL INFORMATION:

, APPLICANT: Wang, David G.  
 , TITLE OF INVENTION: Identification and Mapping of Single Nucleotides  
 , TITLE OF INVENTION: Polymorphisms in the Human Genome  
 , FILE REFERENCE: 108827.129  
 , CURRENT APPLICATION NUMBER: US/10/027,632  
 , CURRENT FILING DATE: 2002-04-30  
 , PRIOR APPLICATION NUMBER: US 60/218,006  
 , PRIOR FILING DATE: 2000-07-12  
 , PRIOR APPLICATION NUMBER: US 60/198,676  
 , PRIOR FILING DATE: 2000-04-20  
 , PRIOR APPLICATION NUMBER: US 60/193,483

PRIOR FILING DATE: 2000-03-29

PRIOR APPLICATION NUMBER: US 60/185,218

PRIOR FILING DATE: 2000-02-24  
PRIOR APPLICATION NUMBER: US

PRIOR FILING DATE: 1999-11-23

PRIOR APPLICATION NUMBER: US

PRIOR FILING DATE: 1999-09-28  
PRIOR APPLICATION NUMBER: IIS 60/145 002

PRIOR FILING DATE: 1999-08-09

NUMBER OF SEQ ID NOS: 325720

SOFTWARE: FastSeq for Windows  
 SEQ ID NO 287049

LENGTH: 600

TYPE: DNA

ORGANISM: Human  
10-027-632-287048

540/97-769-120-01-

Query Match

Best Local Similarity Matches E3: Correlation

March 23; Col

1 CACAAAGTATGGAAATAGAGCTCTGCCCCACACAGTGTGGTACCTTACGGGTTAAGACAGTA 60  
163 CACAACTTGGATATAAATCTGCTC--ATAGTGTGTAATTAGGGTTTAAAAACAGTA 106

```

61 CAGATC 67
    |||||
105 TGATC 99

```

SULT 4  
-10-027-632-287049/c

-10-02

Sequence 287049, Application US/10027632  
Publication No. US20030204075A9  
GENERAL INFORMATION:

TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
 Polymorphisms in the Human Genome  
 TITLE OF INVENTOR: Wang, David S.  
 FILE REFERENCE: 108827.129  
 CURRENT APPLICATION NUMBER: US/10/027, 632  
 CURRENT FILING DATE: 2002-04-30  
 PRIOR APPLICATION NUMBER: US 60/218, 006  
 PRIOR FILING DATE: 2000-07-12  
 PRIOR APPLICATION NUMBER: US 60/198, 676  
 PRIOR FILING DATE: 2000-04-20  
 PRIOR APPLICATION NUMBER: US 60/193, 483  
 PRIOR FILING DATE: 2000-03-29  
 PRIOR APPLICATION NUMBER: US 60/185, 218  
 PRIOR FILING DATE: 2000-02-24  
 PRIOR APPLICATION NUMBER: US 60/167, 363  
 PRIOR FILING DATE: 1999-11-23  
 PRIOR APPLICATION NUMBER: US 60/156, 358  
 PRIOR FILING DATE: 1999-09-28  
 PRIOR APPLICATION NUMBER: US 60/146, 002  
 PRIOR FILING DATE: 1999-08-09  
 NUMBER OF SEQ ID NOS: 325720  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO 287049

TYPE: DNA  
ORGANISM: Human  
10-027-632-287049

Query Match	50.4%;	Score 33.8;	DB 15;	Length 600;
Local Similarity	79.1%;	Pred. No. 0.013;		
Matches	53;	Conservative 0;	Mismatches 12;	Indels 2;
				Gaps 1
1	CACACGCTTGCATAATAAGCTTGC	CCCCACCACTTGGTAACTTTAGGGTTTAAGACAGTA	60	
163	CACAACTTGACATTAATAATCTGCTC	-ATAGTGGTGAATTAAGGTTTAAACAGTA	106	
61	CAGATC	67		

61 CAAGATC 67  
|||||

Db 105 TGAGATC 99

RESULT 5  
US-10-723-860-1311; Sequence 1311, Application US/10723860  
; Publication No. US20040253606A1

; GENERAL INFORMATION:

; APPLICANT: Aziz, Natscha

; APPLICANT: Ginsburg, Wendy M.

; APPLICANT: Zlotnick, Albert

; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &  
; TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators

; FILE REFERENCE: 05882.0193.NPUS01

; CURRENT APPLICATION NUMBER: US/10/723,860

; CURRENT FILING DATE: 2003-11-26

; PRIOR APPLICATION NUMBER: 60/429,739

; PRIOR FILING DATE: 2002-11-26

; NUMBER OF SEQ ID NOS: 8393

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 1311

; LENGTH: 259202

; TYPE: DNA

; ORGANISM: Homo sapiens

US-10-723-860-1311

Query Match 50.4%; Score 33.8; DB 18; Length 259202;  
Best Local Similarity 79.1%; Pred. No. 0.089;  
Matches 53; Conservative 0; Mismatches 12; Indels 2; Gaps 1;

QY 1 CACAGCTTGGAATATAAGTCGCCCCACAGTTGTAACCTTAGGGTTTAAGACAGTA 60

Db 65966 CACAACTTGACATATAATCTCTC-ATAGTGTGTAATTAGGGTTTAAGACAGTA 70023

QY 61 CAAGATC 67

Db 70024 TGAGATC 70030

RESULT 6

US-10-066-543-2458/c

; Sequence 2458, Application US/10066543

; Publication No. US20030087818A1

; GENERAL INFORMATION:

; APPLICANT: Jiang, Yugu

; APPLICANT: Pyle, Ruth A.

; APPLICANT: Xu, Jiangchun

; APPLICANT: Indrtae, Carol Joseph

; APPLICANT: Lodes, Michael J.

; APPLICANT: Secrist, Heather

; APPLICANT: Carter, Derrick

; APPLICANT: Fanger, Gary R.

; APPLICANT: Smith, Carole L.

; APPLICANT: Durham, Margerita

; APPLICANT: Stolk, John A.

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY

; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER

; FILE REFERENCE: 210121.563

; CURRENT APPLICATION NUMBER: US/10/066,543

; CURRENT FILING DATE: 2002-01-31

; NUMBER OF SEQ ID NOS: 3417

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 2458

; LENGTH: 216

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: 150\_202

; OTHER INFORMATION: n = A,T,C or G

US-10-066-543-2458

Query Match 39.7%; Score 26.6; DB 14; Length 216;

Best Local Similarity 63.1%; Pred. No. 5.3;  
Matches 41; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 3 CAAGCTTGGAATATAAGTCGCCCCACAGTTGTAACCTTAGGGTTTAAGACAGTACA 62

Db 81 CAAATATGATATAATATGTTGTGCCCCCAGTTGTTATCTCTAAGATATAAGATATCA 22

QY 63 AGATC 67

Db 21 TGATC 17

RESULT 7

US-10-060-036-3636/c

; Sequence 3636, Application US/10060036

; Publication No. US20030073144A1

; GENERAL INFORMATION:

; APPLICANT: Benson, Darin R.

; APPLICANT: Kalos, Michael D.

; APPLICANT: Persing, David H.

; APPLICANT: Lodes, Michael J.

; APPLICANT: Hepler, William T.

; APPLICANT: Jiang, Yugu

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY

; TITLE OF INVENTION: AND DIAGNOSIS OF PANCREATIC CANCER

; FILE REFERENCE: 210121.566

; CURRENT APPLICATION NUMBER: US/10/060,036

; CURRENT FILING DATE: 2002-01-30

; NUMBER OF SEQ ID NOS: 4560

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 3636

; LENGTH: 559

; TYPE: DNA

; ORGANISM: Homo sapiens

US-10-060-036-3636

Query Match 39.7%; Score 26.6; DB 14; Length 559;  
Best Local Similarity 63.1%; Pred. No. 7.2;  
Matches 41; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 3 CACAGCTTGGAATATAAGTCGCCCCACAGTTGTAACCTTAGGGTTTAAGACAGTACA 62

Db 293 CAAATATGATATAATATGTTGTGCCCCCAGTTGTTATCTCTAAGATATAAGATATCA 234

QY 63 AGATC 67

Db 233 TGATC 229

RESULT 8

US-09-969-034-4304

; Sequence 4304, Application US/09969034

; Publication No. US20040110668A1

; GENERAL INFORMATION:

; APPLICANT: Burgess, Christopher C.

; APPLICANT: Ascle, Jon H.

; APPLICANT: Carroll, Eddie III

; APPLICANT: Catino, Theodore J.

; APPLICANT: Dwivedi, Pooranma

; APPLICANT: Molino, Gary A.

; APPLICANT: Thiagalingam, Arunthathi

; APPLICANT: Lewis, Marcia E.

; TITLE OF INVENTION: Nucleic Acid Sequences Differentially

; TITLE OF INVENTION: Expressed in Cancer Tissue

; FILE REFERENCE: 1657/1032

; CURRENT APPLICATION NUMBER: US/09/969,034

; CURRENT FILING DATE: 2001-10-02

; PRIOR APPLICATION NUMBER: 60/237,271

; PRIOR FILING DATE: 2000-02-10

; NUMBER OF SEQ ID NOS: 4494

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 4304

; LENGTH: 614

TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 225, 481, 514, 524, 599  
OTHER INFORMATION: n = A, T, C or G  
US-09-969-034-4304

Query Match 39.7%; Score 26.6; DB 11; Length 614;  
Best Local Similarity 63.1%; Pred. No. 7.4;  
Matches 41; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 3 CAAGCTTGGAATAAAGTCTGCCCCCAGCTTGTAACCTTTAGGTTTAAGACAGTACA 62  
DB 261 CAATAATGATATATATATGTTTGCCCCCAGTTGTTATCTTAAGATTAAGATTAATCAA 320

QY 63 AGATC 67  
DB 321 TGATC 325

## RESULT 9

US-10-723-860-5715/c  
Sequence 5715, Application US/10723860  
Publication No. US20040253606A1  
GENERAL INFORMATION:  
APPLICANT: Aziz, Natasha  
APPLICANT: Ginsburg, Wendy M.  
APPLICANT: Zlotnik, Albert  
TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &  
TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators  
FILE REFERENCE: 05882.0193.NPUS01  
CURRENT APPLICATION NUMBER: US/10/723,860  
PRIOR FILING DATE: 2003-11-26  
PRIOR APPLICATION NUMBER: 60/429,739  
NUMBER OF SEQ ID NOS: 8393  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 5715  
LENGTH: 2516  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (475)..(489)  
OTHER INFORMATION: n is a, c, g, or t  
US-10-723-860-5715

Query Match 39.7%; Score 26.6; DB 18; Length 2516;  
Best Local Similarity 63.1%; Pred. No. 12;  
Matches 41; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 3 CAAGCTTGGAATAAAGTCTGCCCCCAGCTTGTAACCTTTAGGTTTAAGACAGTACA 62  
DB 1665 CAATAATGATATATATATGTTTGCCCCCAGTTGTTATCTTAAGATTAAGATTAATCAA 1606

QY 63 AGATC 67  
DB 1605 TGATC 1601

## RESULT 10

US-10-723-860-7293/c  
Sequence 7293, Application US/10723860  
Publication No. US20040253606A1  
GENERAL INFORMATION:  
APPLICANT: Aziz, Natasha  
APPLICANT: Ginsburg, Wendy M.  
APPLICANT: Zlotnik, Albert  
TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &  
TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators  
FILE REFERENCE: 05882.0193.NPUS01  
CURRENT APPLICATION NUMBER: US/10/723,860

CURRENT FILING DATE: 2003-11-26  
PRIOR APPLICATION NUMBER: 60/429,739  
PRIOR FILING DATE: 2002-11-26  
NUMBER OF SEQ ID NOS: 8393  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 7293  
LENGTH: 2516  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (475)..(489)  
OTHER INFORMATION: n is a, c, g, or t  
US-10-723-860-7293

Query Match 39.7%; Score 26.6; DB 18; Length 2516;  
Best Local Similarity 63.1%; Pred. No. 12;  
Matches 41; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 3 CAAGCTTGGAATAAAGTCTGCCCCCAGCTTGTAACCTTTAGGTTTAAGACAGTACA 62  
DB 1665 CAATAATGATATATATATGTTTGCCCCCAGTTGTTATCTTAAGATTAAGATTAATCAA 1606

QY 63 AGATC 67  
DB 1605 TGATC 1601

## RESULT 11

US-10-084-817-95/c  
Sequence 95, Application US/10084817  
Publication No. US20030119009A1  
GENERAL INFORMATION:  
APPLICANT: Susan Stuart  
APPLICANT: Jed G. Nuchtern  
APPLICANT: Sharon E. Plon  
APPLICANT: Jason M. Shohet  
TITLE OF INVENTION: GENES REGULATED BY MYCN ACTIVATION  
FILE REFERENCE: PA-0046 US  
CURRENT APPLICATION NUMBER: US/10/084,817  
PRIOR FILING DATE: 2002-02-25  
PRIOR APPLICATION NUMBER: 60/270,784  
PRIOR FILING DATE: 2001-02-23  
NUMBER OF SEQ ID NOS: 365  
SOFTWARE: PERL Program  
SEQ ID NO 95  
LENGTH: 9277  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
OTHER INFORMATION: Incyte ID No. US20030119009A1 1384695.102  
NAME/KEY: unsure  
LOCATION: 1941-1942, 1952-1953, 1956, 1958, 1963, 6600-6601, 9271  
OTHER INFORMATION: a, t, c, g, or other  
US-10-084-817-95

Query Match 39.7%; Score 26.6; DB 15; Length 9277;  
Best Local Similarity 63.1%; Pred. No. 18;  
Matches 41; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 3 CAAGCTTGGAATAAAGTCTGCCCCCAGCTTGTAACCTTTAGGTTTAAGACAGTACA 62  
DB 5902 CAATAATGATATATATATGTTTGCCCCCAGTTGTTATCTTAAGATTAAGATTAATCAA 5843

QY 63 AGATC 67  
DB 5842 TGATC 5838

## RESULT 12

US-10-027-632-282653  
Sequence 282653, Application US/10027632

```
/ Publication No. US20020198371A1
/ GENERAL INFORMATION:
/ APPLICANT: Wang, David G.
/ TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
/ FILE REFERENCE: 108827.129
/ CURRENT APPLICATION NUMBER: US/10/027,632
/ PRIOR FILING DATE: 2002-04-30
/ PRIOR APPLICATION NUMBER: US 60/218,006
/ PRIOR FILING DATE: 2000-07-12
/ PRIOR APPLICATION NUMBER: US 60/198,676
/ PRIOR FILING DATE: 2000-04-20
/ PRIOR APPLICATION NUMBER: US 60/193,483
/ PRIOR FILING DATE: 2000-03-29
/ PRIOR APPLICATION NUMBER: US 60/185,218
/ PRIOR FILING DATE: 2000-02-24
/ PRIOR APPLICATION NUMBER: US 60/167,363
/ PRIOR FILING DATE: 1999-11-23
/ PRIOR APPLICATION NUMBER: US 60/156,358
/ PRIOR FILING DATE: 1999-09-28
/ PRIOR APPLICATION NUMBER: US 60/146,002
/ PRIOR FILING DATE: 1999-08-09
/ NUMBER OF SEQ ID NOS: 325720
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 282653
/ LENGTH: 617
/ TYPE: DNA
/ ORGANISM: Human
US-10-027-632-282653

Query Match          39.4%; Score 26.4; DB 13; Length 617;
Best Local Similarity 65.0%; Pred. No. 8.9;
Matches 39; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY      8 TTGGAATTAAGTCTGCCCCACCAGTTGTAAGGTTTAAAGACAGTCAAGATC 67
Db      184 TTGGAGTAACAGTCTACCCCAAACTTAGTACCTCATGTATTAACATTCAAAAAC 243

RESULT 13
US-10-027-632-282654
/ Sequence 282654, Application US/10027632
/ Publication No. US20020198371A1
/ GENERAL INFORMATION:
/ APPLICANT: Wang, David G.
/ TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
/ FILE REFERENCE: 108827.129
/ CURRENT APPLICATION NUMBER: US/10/027,632
/ PRIOR FILING DATE: 2002-04-30
/ PRIOR APPLICATION NUMBER: US 60/218,006
/ PRIOR FILING DATE: 2000-07-12
/ PRIOR APPLICATION NUMBER: US 60/198,676
/ PRIOR FILING DATE: 2000-04-20
/ PRIOR APPLICATION NUMBER: US 60/193,483
/ PRIOR FILING DATE: 2000-03-29
/ PRIOR APPLICATION NUMBER: US 60/185,218
/ PRIOR FILING DATE: 2000-02-24
/ PRIOR APPLICATION NUMBER: US 60/167,363
/ PRIOR FILING DATE: 1999-11-23
/ PRIOR APPLICATION NUMBER: US 60/156,358
/ PRIOR FILING DATE: 1999-09-28
/ PRIOR APPLICATION NUMBER: US 60/146,002
/ PRIOR FILING DATE: 1999-08-09
/ NUMBER OF SEQ ID NOS: 325720
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 282654
/ LENGTH: 617
/ TYPE: DNA
/ ORGANISM: Human
US-10-027-632-282654

Query Match          39.4%; Score 26.4; DB 13; Length 617;
Best Local Similarity 65.0%; Pred. No. 8.9;
Matches 39; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
```

```
Best Local Similarity 65.0%; Pred. No. 8.9;
Matches 39; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY      8 TTGGAATTAAGTCTGCCCCACCAGTTGTAAGGTTTAAAGACAGTCAAGATC 67
Db      184 TTGGAGTAACAGTCTACCCCAAACTTAGTACCTCATGTATTAACATTCAAAAAC 243

RESULT 14
US-10-027-632-282653
/ Sequence 282653, Application US/10027632
/ Publication No. US20030204075A9
/ GENERAL INFORMATION:
/ APPLICANT: Wang, David G.
/ TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
/ FILE REFERENCE: 108827.129
/ CURRENT APPLICATION NUMBER: US/10/027,632
/ PRIOR FILING DATE: 2002-04-30
/ PRIOR APPLICATION NUMBER: US 60/218,006
/ PRIOR FILING DATE: 2000-07-12
/ PRIOR APPLICATION NUMBER: US 60/198,676
/ PRIOR FILING DATE: 2000-04-20
/ PRIOR APPLICATION NUMBER: US 60/193,483
/ PRIOR FILING DATE: 2000-03-29
/ PRIOR APPLICATION NUMBER: US 60/185,218
/ PRIOR FILING DATE: 2000-02-24
/ PRIOR APPLICATION NUMBER: US 60/167,363
/ PRIOR FILING DATE: 1999-11-23
/ PRIOR APPLICATION NUMBER: US 60/156,358
/ PRIOR FILING DATE: 1999-09-28
/ PRIOR APPLICATION NUMBER: US 60/146,002
/ PRIOR FILING DATE: 1999-08-09
/ NUMBER OF SEQ ID NOS: 325720
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 282653
/ LENGTH: 617
/ TYPE: DNA
/ ORGANISM: Human
US-10-027-632-282653

Query Match          39.4%; Score 26.4; DB 15; Length 617;
Best Local Similarity 65.0%; Pred. No. 8.9;
Matches 39; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY      8 TTGGAATTAAGTCTGCCCCACCAGTTGTAAGGTTTAAAGACAGTCAAGATC 67
Db      184 TTGGAGTAACAGTCTACCCCAAACTTAGTACCTCATGTATTAACATTCAAAAAC 243

RESULT 15
US-10-027-632-282654
/ Sequence 282654, Application US/10027632
/ Publication No. US20030204075A9
/ GENERAL INFORMATION:
/ APPLICANT: Wang, David G.
/ TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
/ FILE REFERENCE: 108827.129
/ CURRENT APPLICATION NUMBER: US/10/027,632
/ PRIOR FILING DATE: 2002-04-30
/ PRIOR APPLICATION NUMBER: US 60/218,006
/ PRIOR FILING DATE: 2000-07-12
/ PRIOR APPLICATION NUMBER: US 60/198,676
/ PRIOR FILING DATE: 2000-04-20
/ PRIOR APPLICATION NUMBER: US 60/193,483
/ PRIOR FILING DATE: 2000-03-29
/ PRIOR APPLICATION NUMBER: US 60/185,218
/ PRIOR FILING DATE: 2000-02-24
/ PRIOR APPLICATION NUMBER: US 60/167,363
/ PRIOR FILING DATE: 1999-11-23
/ PRIOR APPLICATION NUMBER: US 60/156,358
/ PRIOR FILING DATE: 1999-09-28
```

; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 282654  
; LENGTH: 617  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-282654

Query Match 39.4%; Score 26.4; DB 15; Length 617;  
Best Local Similarity 65.0%; Pred. No. 8.9;  
Matches 39; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

Qy 8 TTGGAATTAAGTCTGCCCCCACCAGTTGTTAACTTTAGGGTTTAAGACAGTACAAGATC 67  
Db 184 TTGGAGTACAGTCTACCCCAACTTAGTAGCTTCATGTATTAACATTTCATATAAAC 243

Search completed: February 9, 2005, 06:51:04  
Job time : 93.2823 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 8, 2005, 12:58:25 ; Search time 798.033 Seconds  
(without alignments)  
3059.349 Million cell updates/sec

Title: US-10-754-437-8

Perfect score: 67  
Sequence: 1 cacaagcttggaataaag.....gttaagacagcacaagatc 67

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :  
1: gb\_esc1:\*  
2: gb\_esc2:\*  
3: gb\_esc3:\*  
4: gb\_esc4:\*  
5: gb\_esc5:\*  
6: gb\_esc6:\*  
7: gb\_esc7:\*  
8: gb\_esc8:\*  
9: gb\_esc9:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	31	46.3	513	8	AZ725508 RPCI-24-1
2	31	46.3	513	8	BH118642 RPCI-24-3
3	30.2	45.1	569	2	BF054120 EST439350
4	30.2	45.1	703	5	BO509607 EST617022
5	29.2	43.6	802	9	CL315981 ZMWB0050
6	28.8	43.0	670	9	CE826965 tigr-g88-
7	28.6	42.7	584	2	BE463065 EST343360
8	28.2	42.1	630	2	BB660453 BB660453
9	28.2	42.1	708	8	AZ263719 RPCI-23-1
10	28.2	42.1	729	9	AG505646 Mus muscu
11	27.8	41.5	232	2	BB214296 BB214296
12	27.8	41.5	469	8	AZ617508 1M0448N17
13	27.6	41.2	421	2	BB786603 BB786603
14	27.4	40.9	574	7	CN667917 A0860C04-
15	27.4	40.9	587	8	AQ410748 HS_5067_A
16	27.4	40.9	698	9	AG297816 Mus muscu
17	27.2	40.6	598	7	AG295587 Mus muscu
18	27.2	40.6	598	7	CO379630 CO379630
19	27.2	40.6	1002	1	AL576285 AL576285
20	27	40.3	408	8	AQ635311 RPCI-11-4
21	27	40.3	772	9	CL375759 ZMWB0039
22	26.8	40.0	206	1	AV255261 AV255261
23	26.8	40.0	351	1	AJ559892 AJ559892
24	26.8	40.0	413	4	BG958108 CM0-CT080

25	26.8	40.0	672	9	BX147161 BX147161
26	26.8	40.0	701	9	BX185444 BX185444
27	26.6	39.7	294	7	T04932 T04932
28	26.6	39.7	331	1	AA330839 AA330839
29	26.6	39.7	342	1	AA262024 AA262024
30	26.6	39.7	443	1	AA442029 AA442029
31	26.6	39.7	488	6	CA333757 CA333757
32	26.6	39.7	489	8	BZ262614 BZ262614
33	26.6	39.7	498	1	AA176275 AA176275
34	26.6	39.7	499	1	AA176332 AA176332
35	26.6	39.7	540	2	AW952641 AW952641
36	26.6	39.7	572	9	CR333642 CR333642
37	26.6	39.7	584	5	BU072790 BU072790
38	26.6	39.7	635	6	CA948269 CA948269
39	26.6	39.7	641	6	CB127044 CB127044
40	26.6	39.7	699	4	BG623734 BG623734
41	26.6	39.7	713	1	AV717150 AV717150
42	26.6	39.7	737	5	BX411376 BX411376
43	26.6	39.7	777	4	BM724386 BM724386
44	26.6	39.7	797	5	BX473434 BX473434
45	26.6	39.7	832	9	CR293936 CR293936

#### ALIGNMENTS

RESULT 1  
AZ725508 513 bp DNA linear GSS 24-JAN-2001  
RPCI-24-109J1-TV RPCI-24 Mus musculus genomic clone RPCI-24-109J1,  
DEFINITION genomic survey sequence.  
ACCESSION AZ725508  
VERSION AZ725508.1 GI:12472231  
KEYWORDS GSS.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akintrel,B., Leavins,M., Tsengaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregiorgis,E., Russell,D., de Jong,P. and Fraser,C.M.  
Mouse BAC End Sequences from Library RPCI-24  
Unpublished (1999)  
Contact: Shaying Zhao  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-24. For BAC library availability, please contact Pieter de Jong (pdjong@mail.cho.org). Clones may be purchased from BACPAC Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end page: [http://www.tigr.org/tdb/bac\\_ends/mouse/bac\\_end\\_intro.html](http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html)  
Plate: 109 row: J column: 1  
Seq primer: T7  
Class: BAC ends.

#### FEATURES

source Location/Qualifiers  
1..513  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="RPCI-24-109J1"  
/sex="Male"  
/cell\_type="Spleen/Brain"  
/clone\_lib="RPCI-24"  
/note="Vector: pTRABAC1, Site 1: BamHI, Site 2: BamHI, RPCI-24 Mouse BAC Library produced by Pieter de Jong. The library was cloned in the pTRABAC1 cloning vector at the BamHI sites using MboI partially digested male C57BL/6J





KEYWORDS EST.  
SOURCE Solanum tuberosum (potato)  
ORGANISM Solanum tuberosum  
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; lamiales; Solanales; Solanaceae; Solanum.  
AUTHORS 1 (bases 1 to 703)  
Buell, C.R., Hart, A., Baker, B., Tanksey, S., Fry, W., Smart, C., Restrepo, S., Griffiths, H., van der Hoeven, K., Tsai, U. and Karymcheva, S.A.  
TITLE Generation of a set of potato cDNA clones for microarray analyses  
JOURNAL Unpublished (2002)  
COMMENT On Jun 10, 2002 this sequence version replaced gi:21368476.  
Other ESTs: EST617020 EST617021  
Contact: Rodin Buell  
The Institute for Genomic Research  
9112 Medical Center Dr, Rockville, MD 20850, USA  
Email: potatoc-array@tigr.org  
This clone can be obtained from the University of Arizona Genomics Institute. Orders can be made through URL: <http://genome.arizona.edu/orders/>  
Seq primer: T7.  
Location/Qualifiers  
1..703  
/organism="Solanum tuberosum"  
/mol\_type="mRNA"  
/cultivar="Kennebec or Binjite"  
/db\_xref="taxon:4113"  
/clone="STM93"  
/issue\_type="mixed tissues"  
/lab\_host="SOLR"  
/clone\_lib="generation of a set of potato cDNA clones for microarray analyses mixed potato tissues"  
/note="Vector: Bluescript SK(-); Site 1: EcoRI; Site 2: XhoI; supplier: Combination of untreated and Phytophthora infestans-treated libraries of stolons, leaves, leaflets, axillary buds of stem explants, petioles, germinating eyes, tubers, or roots."

ORIGIN  
Query Match 45.1%; Score 30.2; DB 5; Length 703;  
Best Local Similarity 65.7%; Pred. No. 4.8;  
Matches 44; Conservative 0; Mismatches 23; Indels 0; Gaps 0;  
QY 1 CACAACTTGAATAAAGTCTGCCACCAAGTGTGAATCTTAGGTTTAAGACAGTA 60  
DB 548 CACATGCTTCATCATTAATTTCTTCAACACCAAGTCTTAAGGTTTGTGACATGA 489  
QY 61 CAAATC 67  
DB 488 CAACTTC 482

RESULT 5  
CLJ15981 802 bp DNA linear GSS 03-MAR-2004  
LOCUS ZMBC0502B08 ZMBC0502B08 Zea mays genomic clone ZMBC0502B08 3',  
DEFINITION genomic survey sequence.  
ACCESSION CLJ15981  
VERSION CLJ15981.1 GI:44895870  
KEYWORDS GSS.  
SOURCE Zea mays  
ORGANISM Zea mays  
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.  
AUTHORS 1 (bases 1 to 802)  
Bharti, A.K., Young, S., Kavchok, S., Kiefer, G., Bronzino, A.C., Zohovetz, V., Fuks, G., Yu, Y., Wing, R., and Messing, J.  
TITLE Sequencing of the maize genome at FeIR (2003c)  
JOURNAL Unpublished (2003)  
COMMENT Contact: Bharti, A.K.  
Dr. Joachim Messing's lab

THE Plant Genome Initiative at Rutgers, Wakeman Institute, Rutgers University  
190 Frelinghuysen Road, Piscataway, NJ 08854, USA  
Tel: 732 445 3801  
Fax: 732 445 5735  
Email: bharti@wakeman.rutgers.edu  
Seq primer: SP6  
Classes: BAC ends.  
Location/Qualifiers  
1..802  
/organism="Zea mays"  
/mol\_type="genomic DNA"  
/cultivar="B73"  
/db\_xref="taxon:4577"  
/clone="ZMBC0502B08"  
/lab\_host="E. coli DH10B"  
/clone\_lib="ZMBC"  
/note="Vector: pTARBAC1.3; Site\_1: BamHI; Site\_2: BamHI"

ORIGIN  
Query Match 43.6%; Score 29.2; DB 9; Length 802;  
Best Local Similarity 69.0%; Pred. No. 11;  
Matches 40; Conservative 0; Mismatches 18; Indels 0; Gaps 0;  
QY 5 AGCTTGAATAAAGTCTGCCACCAAGTGTGAATCTTAGGTTTAAGACAGTA 62  
DB 530 AGATTGAACCTCAAGTGAACACCCAGGTTAGCTTAAGGTTTAAGTCAACACA 587

RESULT 6  
CE826965 670 bp DNA linear GSS 30-SEP-2003  
LOCUS tigr-gss-dog-17000331964321 Dog Library Canis familiaris genomic,  
DEFINITION genomic survey sequence.  
ACCESSION CE826965  
VERSION CE826965.1 GI:37167985  
KEYWORDS GSS.  
SOURCE Canis familiaris (dog)  
ORGANISM Canis familiaris  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
AUTHORS 1 (bases 1 to 670)  
Kirkness, E.F., Batfa, V., Halpern, A.L., Levy, S., Remington, K., Ruesch, D.B., Delcher, A.L., Pop, M., Wang, W., Frazer, C.M. and Venter, J.C.  
TITLE The dog genome: survey sequencing and comparative analysis  
JOURNAL Science 301 (5641), 1898-1903 (2003)  
MEDLINE 22875432  
PUBMED 14512627  
COMMENT Contact: Kirkness EF  
The Institute for Genomic Research  
Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,  
Rockville, MD 20850, USA  
Tel: 301-838-0200  
Fax: 301-838-0208  
Email: ekirknes@tigr.org  
Class: shotgun.  
Location/Qualifiers  
1..670  
/organism="Canis familiaris"  
/mol\_type="genomic DNA"  
/strain="Standard Poodle"  
/db\_xref="taxon:9615"  
/clone\_lib="Dog Library"  
/note="Site 1: BstXI; Libraries were prepared from peripheral blood"

ORIGIN  
Query Match 43.0%; Score 28.8; DB 9; Length 670;  
Best Local Similarity 65.6%; Pred. No. 15;  
Matches 42; Conservative 0; Mismatches 22; Indels 0; Gaps 0;  
QY 1 CACAACTTGAATAAAGTCTGCCACCAAGTGTGAATCTTAGGTTTAAGACAGTA 60

DB 478 CAAAAGCTGAGAAATTAATTCATGCTCATCCCATGCAAGCTTTATGTTTAAAGAGATTA 419  
 QY 61 CAG 64  
 DB 418 AAG 415

RESULT 7  
 BE463065/c 584 bp mRNA linear EST 18-MAY-2001  
 LOCUS EST354360 tomato flower buds 8 mm to pre-anthesis, Cornell  
 DEFINITION University Lycopersicon esculentum cDNA clone CTC011F24, mRNA  
 sequence  
 BE463065  
 ACCESSION BE463065.1 GI:9508908  
 VERSION  
 KEYWORDS  
 SOURCE Lycopersicon esculentum (tomato)  
 ORGANISM Lycopersicon esculentum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 asterids; lamiales; Solanales; Solanaceae; Solanum; Lycopersicon.  
 1 (bases 1 to 584)  
 van der Hoeven, R.S., Bezzerides, J.L., Matern, A.L., Holt, I.E.,  
 Liang, F., Upton, J., Hansen, T., Craven, M.B., Bowman, C.L., Ahn, S.,  
 Roming, C.M., Fraser, C.M., Martin, G.B., Giovannoni, J.J. and  
 Tanksley, S.D.  
 Generation of ESTs from tomato flower tissue  
 Title Unpublished (1999)  
 JOURNAL  
 COMMENT Contact: CUGI  
 Clemson University  
 100 Jordan Hall, Clemson, SC 29634, USA  
 Email: <http://www.genome.clemson.edu/orders/index.html>  
 5 prime sequence.

FEATURES  
 source  
 1..584  
 /organism="Lycopersicon esculentum"  
 /mol\_type="mRNA"  
 /cui\_livar="TA496"  
 /db\_xref="taxon:4081"  
 /clone="CTC011F24"  
 /tissue\_type="flower"  
 /dev\_stage="buds 8mm-to-preanthesis"  
 /clone\_lib="tomato flower buds 8 mm to pre-anthesis,  
 Cornell University"  
 /note="Vector: pBluescript SK(-); Site\_1: EcoRI; Site\_2:  
 XhoI; supplier: Tanksley; Flower buds and flowers were  
 taken from greenhouse plants (4-8 wks old, TA496). They  
 were immediately frozen in liquid nitrogen and then  
 size-separated while remaining frozen."

ORIGIN  
 Query Match 42.7%; Score 28.6; DB 2; Length 584;  
 Best Local Similarity 64.2%; Pred. No. 18;  
 Matches 43; Conservative 0; Mismatches 24; Indels 0; Gaps 0;  
 QY 1 CACAAAGCTTGAATAAAGTCGCGCCACGAGTTGTAAGCTTTAAGACAGTA 60  
 DB 543 CACATGCTTCATCATTAATTTCTTCAACACCAAGCTTTAAGGCAAGGTTTGAACATGA 484  
 QY 61 CAAAGATC 67  
 DB 483 CAAAGTTC 477

RESULT 8  
 BB660453/c 630 bp mRNA linear EST 26-OCT-2001  
 LOCUS BB660453 RIKEN full-length enriched, 13 days embryo lung Mus  
 DEFINITION musculus cDNA clone D430047118 5', mRNA sequence.  
 ACCESSION BB660453  
 VERSION BB660453.1 GI:16494274

KEYWORDS  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 630)

REFERENCE  
 Arikawa, T., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T.,  
 Hara, A., Hiramoto, K., Hori, F., Ishii, Y., Ito, M., Kawai, J.,  
 Kono, H., Kouda, M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K.,  
 Ono, M., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K.,  
 Sano, H., Sasaki, D., Shibata, K., Shingawa, A., Shiraki, T.,  
 Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, P.,  
 Takeeda, Y., Tanaka, T., Toyota, T., Muramatsu, M. and Hayashizaki, Y.  
 RIKEN Mouse ESTs (Arikawa, T., et al. 2001)  
 Title Unpublished (2001)  
 JOURNAL  
 COMMENT Contact: Yoshihide Hayashizaki  
 Laboratory for Genome Exploration Research Group, RIKEN Genomic  
 Sciences Center (GSC), Yokohama Institute  
 The Institute of Physical and Chemical Research (RIKEN)  
 1-7-22 Suenho-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 Tel: 81-45-503-9222  
 Fax: 81-45-503-9216  
 Email: [genome-research.riken.jp](mailto:genome-research.riken.jp), URL: <http://genome.gsc.riken.jp/>  
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
 Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
 Normalization and subtraction of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new  
 genes. Genome Res. 10 (10), 1617-1630 (2000)  
 wagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,  
 Watanishi, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T.,  
 Matsuzawa, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A.  
 and Hayashizaki, Y.  
 RIKEN integrated sequence analysis (RISA) system--384-format  
 sequencing pipeline with 384 multichannel sequencer. Genome Res.  
 10 (11), 1757-1771 (2000)  
 Kono, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P.,  
 Sugahara, Y. and Hayashizaki, Y.  
 Computer-based methods for the mouse full-length cDNA  
 encyclopedia: real-time sequence clustering for construction of a  
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
 Aizawa, K., Shinagawa, A., Saito, T., Kiyosawa, H., Yamashita, I.,  
 Aizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and  
 Hayashizaki, Y.  
 Computational Analysis of Full-length Mouse cDNAs Compared with  
 Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)  
 Please visit our web site (<http://genome.gsc.riken.go.jp>) for  
 further details.  
 e mouse tissues.

FEATURES  
 source  
 1..630  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /clone="D430047118"  
 /tissue\_type="lung"  
 /dev\_stage="13 days embryo"  
 /lab\_host="MDH10B"  
 /clone\_lib="RIKEN full-length enriched, 13 days embryo  
 lung"  
 /note="Site\_1: SalI; Site\_2: BamHI; cDNA library was  
 prepared and sequenced in Mouse Genome Encyclopedia  
 Project of Genome Exploration Research Group in Riken  
 Genomic Sciences Center and Genome Science Laboratory in  
 RIKEN. Division of Experimental Animal Research in Riken  
 contributed to prepare mouse tissues. 1st strand cDNA was  
 primed with a primer [5',  
 GAGAGAGAGACGCGCCGACCACTGAGTTTCTTTTCTTTTNN 3'], cDNA was  
 prepared by using trehalose thermo-activated reverse  
 transcriptase and subsequently enriched for full-length by  
 cap-trapper. Second strand cDNA was prepared with the  
 primer adapter of sequence [5',  
 GAGAGAGAGATTCTCGAGTTAATTAATTAATCCGCCGCCGCC 3']. cDNA  
 was cleaved with BamHI and XhoI. Vector: a modified

ORIGIN pb1uescript KS(+) after bulk excision from Lambda FLC I."

Query Match 42.1%; Score 28.2; DB 2; Length 630;  
Best Local Similarity 64.6%; Pred. No. 25;  
Matches 42; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 2 ACAAGCTTGAATTAAGCTGCCCCACAGTTGTAAGTTAAGCAGTAC 61  
DB 532 AGAACTTAAATATACAACTGCCCCAAAGTAGTACTCTGTTAAGTGTCAATTG 473

QY 62 AAGAT 66  
DB 472 AAGAT 468

RESULT 9 708 bp DNA linear GSS 26-JUL-2000  
A2263719  
LOCUS RPT-23-153M12, TV RPT-23 Mus musculus genomic clone  
DEFINITION RPT-23-153M12, genomic survey sequence.  
ACCESSION A2263719  
VERSION A2263719.1 GI:9474235  
KEYWORDS GSS.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 708)

REFERENCE Zhao, S., Nieman, W., Feldblum, T., Malek, J., Shatsman, S.,  
Akintet, B., Levine, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de  
Mouze, P. and Fraser, C.M. 1999. RPT-23-153M12. TV  
Mouse BAC End Sequences from Library RPT-23  
Unpublished (1999)

COMMENT Other GSSs: RPT-23-153M12.TV  
Contact: Shaying Zhao  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
7112 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPT-23. For BAC  
library availability, please contact Pieter de Jong  
(pieterdejong.med.bufralo.edu). Clones may be purchased from  
BACPAC Resources (<http://bacpac.med.bufralo.edu/orderingframe.html>)  
or from Resea ch Genetics ([http://www.tigr.org/tdb/bac\\_end/mouse/bac\\_end\\_intro.html](http://www.tigr.org/tdb/bac_end/mouse/bac_end_intro.html))  
Plate: 153 row: M column: 12  
Seq primer: T7  
Class: BAC ends.

FEATURES Location/Qualifiers

source 1..708  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="RPT-23-153M12"  
/sex="Female"  
/lab\_host="DH10B"  
/clone\_lib="RPT-23"  
/note="Organ: Kidney/Brain; Vector: pBAC3.6; Site: 1.  
ECORI; Site: 2. EORI; Female C57BL/6J mouse kidney and/or  
brain genomic DNA was isolated and partially digested  
with a combination of EORI and EORI Methylase. Size  
selected DNA was cloned into the pBAC3.6 vector at the  
ECORI sites. The ligation products were transformed into  
DH10B electrocompetent cells (BRL Life Technologies)."

ORIGIN

Query Match 42.1%; Score 28.2; DB 8; Length 708;  
Best Local Similarity 64.6%; Pred. No. 26;  
Matches 42; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 2 ACAAGCTTGAATTAAGCTGCCCCACAGTTGTAAGTTAAGCAGTAC 61  
DB 495 AGAACTTAAATATACAACTGCCCCAAAGTAGTACTCTGTTAAGTGTCAATTG 436  
QY 62 AAGAT 66  
DB 435 AAGAT 431

RESULT 10 729 bp DNA linear GSS 04-JUN-2004  
AG505646  
LOCUS Mus musculus molossinus DNA, clone:MSWg01-407M07.T7, genomic survey  
sequence.  
ACCESSION AG505646  
VERSION AG505646.1 GI:48213059  
KEYWORDS GSS.  
SOURCE Mus musculus molossinus  
ORGANISM Mus musculus molossinus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1  
Hattori, M., Toyoda, A., Noguchi, H., Kojima, T. and Sakaki, Y.  
BAC end Sequences of Library MSWg01  
Unpublished  
2 (bases 1 to 729)  
Hattori, M., Toyoda, A., Noguchi, H., Kojima, T. and Sakaki, Y.  
Direct Submission  
Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical  
and Chemical Research (RIKEN), Genomic Sciences Center (GSC),  
1-7-22 Suenho-chu, Tsukuba, Ibaraki, Japan, Yokohama, Kanagawa 230-0045, Japan  
(E-mail: hattori@gsc.riken.jp, URL: <http://hgp.gsc.riken.go.jp/>,  
Tel: 81-45-503-9111, Fax: 81-45-503-9170)  
Clones are derived from the mouse BAC library MSWg01. For BAC  
library availability, please contact Kunya Abe (abe@rc.riken.jp).  
Tsukuba Institute, Bio Resource Center.  
The Institute of Physical and Chemical Research (RIKEN) 3-1-1  
Koyada, Tsukuba, 305-0074 Japan  
phone: 81-298-36-9189, fax: 81-298-36-9199  
e-mail: abe@rc.riken.jp

COMMENT

LIBRARY : T7  
Vector : pBAC3.6  
R.Site 1 : EORI  
R.Site 2 : EORI.  
Sequencing : T7  
Location/Qualifiers

source 1..729  
/organism="Mus musculus molossinus"  
/mol\_type="genomic DNA"  
/sub\_species="molossinus"  
/db\_xref="taxon:57486"  
/clone="MSWg01-407M07.T7"  
/sex="male"  
/issue\_type="mixture of kidney and spleen"  
/clone\_lib="MSWg01 Mouse Male BAC Library"

ORIGIN

Query Match 42.1%; Score 28.2; DB 9; Length 729;  
Best Local Similarity 64.6%; Pred. No. 26;  
Matches 42; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 2 ACAAGCTTGAATTAAGCTGCCCCACAGTTGTAAGTTAAGCAGTAC 61  
DB 592 AGAACTTAAATATACAACTGCCCCAAAGTAGTACTCTGTTAAGTGTCAATTG 651

QY 62 AAGAT 66  
DB 652 AAGAT 656

RESULT 11

[illegible]

to Rot = 20.0 and subtraction to Rot = 459.0. Second strand cDNA was prepared with the primer adapter of sequence [5'] GACGAGAGATTCGACTGATTAATTAATCCCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

ORIGIN

Query Match 41.5%; Score 27.8; DB 2; Length 232;  
Beet Local Similarity 69.1%; Pred. No. 29;  
Matches 38; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

Db 56 CACAGAGTGGGAACTTAAAGGCGCCGCCACACAGGTGGGATATATAGGCATATGA 2

RESULT 12  
A2617508/c

LOCUS  
DEFINITION  
1M0448N1R Mouse 10kb plasmid UGCLM library Mus musculus genomic clone UGCLM0448N17 R, genomic survey sequence.

ACCESSION  
A2617508

VERSION  
A2617508.1 GI:11739698

KEYWORDS  
GSS.

SOURCE  
Mus musculus (house mouse)

ORGANISM  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eumelastomi; Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murine; Mus.

REFERENCE  
1 (bases 1 to 469)  
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausen,A. and Wright,D. Weis,R.  
Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts  
Unpublished (2000)

TITLE  
Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0448 row: N column: 17  
Seq primer: CACACAGGAACAGCATATGACC  
Class: plasmid ends  
High quality sequence stop: 469.

JOURNAL  
COMMENT

FEATURES  
Source  
1. 469  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UGCLM0448N17"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/clone\_lib="Mouse 10kb plasmid UGCLM library"  
/note="Vector: pMD24nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (gi|4732114|gb|AF129072.1), a copy-number

inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

## ORIGIN

Query Match 41.5%; Score 27.8; DB 8; Length 469;  
Best Local Similarity 65.1%; Pred. No. 33;  
Matches 41; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 3 CAAGCTTGAAATGAAGTCTGCCCCACGATTGTAACCTTAGGCTTAAGACAGTACA 62

DB 205 CAGCTTGAGAGATTAAGTCTGCCACCCACACATTAATTATGCTTGAGCCAGTAC 146

QY 63 AGA 65

DB 145 AGA 143

## RESULT 13

BB786603

LOCUS

BB786603 RIKEN full-length cDNA clone G430112H24 3', mRNA sequence.

DEFINITION

musculus cDNA clone G430112H24 3', mRNA sequence.

ACCESSION

BB786603

VERSION

BB786603.1

KEYWORDS

EST.

SOURCE

Mus musculus

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1 (bases 1 to 421)

Akimura, T., Arakawa, T., Carninci, P., Furuno, M., Hanagaki, T., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T., Imotani, K., Ishii, Y., Ito, M., Kawai, J., Kojima, Y., Kono, H., Konda, M., Matsuyama, T., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagawa, A., Takahashi, F., Takaku-kahira, S., Tanaka, T., Tomaru, A., Toyota, T., Watahiki, A., Yasunishi, A., Yamanaka, M., and Hayashizaki, Y.

Muramatsu, M. and Hayashizaki, Y.

RIKEN Encyclopedia of Mouse Full-length cDNAs (AKIMURA, T., et al. 2001)

Unpublished (2001)

Contact: Yoshitake Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suenho-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@sc.riken.jp, URL: http://genome-gsc.riken.jp/

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)

wagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Kono, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y. and Hayashizaki, Y.

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

Please visit our web site (<http://genome-gsc.riken.go.jp>) for further details.

e mouse tissues.

## FEATURES

Source

1. 421

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="BALB/c"

/db\_xref="taxon:10090"

/clone="G430112H24"

/tissue\_type="Blood"

/cell\_line="RCB-0035 WEHI-3"

/clone\_11b="RIKEN full-length enriched, RCB-0035 WEHI-3 cDNA"

ORIGIN

Query Match 41.2%; Score 27.6; DB 2; Length 421;  
Best Local Similarity 67.2%; Pred. No. 38;  
Matches 39; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 10 GGAATTAAGTCTGCCCCACGATTGTAACCTTAGGCTTAAGACAGTACAAGTAC 67

DB 317 GGCAAGATTAAGTCTGCCCCACGATTGTAACCTTAGGCTTAAGACAGTATTTATC 374

## RESULT 14

CN667917/c

LOCUS

CN667917-5 NIA Mouse E13.5 whole embryo cDNA library (Long) Mus

DEFINITION

musculus cDNA clone NIA:A0860C04 IMAGE:30762939 5', mRNA sequence.

ACCESSION

CN667917

VERSION

CN667917.1

KEYWORDS

EST.

SOURCE

Mus musculus

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1 (bases 1 to 574)

Sharov, A.A., Piao, Y., Marob, R., Dudekula, D.B., Qian, Y., Vanburen, V., Falco, G., Martin, P.R., Stragg, C.A., Basse, U.C., Wang, Y., Carter, M.G., Hamatani, T., Alpa, K., Akutsu, H., Sharova, L., Tanaka, T.S., Kimber, W.L., Yoshikawa, T., Jaradat, S.A., Pantano, S., Nagaraja, R., Boheler, K.R., Taub, D., Hodges, R.J., Longo, D.L., Schlessinger, D., Keller, J., Klotz, E., Kelsae, G., Umazawa, A., Vescevi, A.L., Rosant, J., Kunath, T., Hogan, B.L., Curci, A., D'Urso, M., Kelsae, G., Hide, W. and Ko, M.S.

Transcriptome analysis of mouse stem cells and early embryos

Plos Biol. 1 (3), 410-419 (2003)

Contact: Dawood B. Dudekula

Laboratory of Genetics

National Institute on Aging/National Institutes of Health

333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA

Email: [cdna@lgsun.grc.nia.nih.gov](mailto:cdna@lgsun.grc.nia.nih.gov)

Plate: A0860 row: C column: 04

Seq primer: M13 Reverse

High quality sequence stop: 574

POLYVA-No.

FEATURES

Source

1. 574

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="NIA:A0860C04 IMAGE:30762939"

/tissue\_type="whole embryo including extraembryonic tissues at 13.5-days postcoitum"

/dev stage="E13.5"

/lab host="DH10B"

/clone\_11b="NIA Mouse E13.5 whole embryo cDNA library (Long)"

/note="Vector: pCMV-SPORT6 (Invitrogen); Site 1: SalI; Site 2: NotI; Mouse cDNA project by the Laboratory of Genetics, National Institute on Aging (NIA), Intramural Research Program, NIH (<http://lgsun.grc.nia.nih.gov/cdna>).

This is a long-transcript enriched cDNA library (Ref. Genome Res. 11: 1553-1558 (2001). [PMID: 11544191]). Total RNAs were extracted from 1 embryo at 13.5-days postcoitum. Double-stranded cDNAs were synthesized with an oligo(dT) primer [Invitrogen: 5'-pGACTGATCTTAAATCGCGAGCGCCCTTTTCTTTT-3'] from 3ug of total RNA, treated with T4 DNA polymerase, and purified by ethanol-precipitation. The cDNAs were ligated to lone-linker L1-Sal1, purified by phenol/chloroform, and separated from free linkers by Centricon 100. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Taq polymerase (Takara) with a primer Sal1-S. The products were purified by phenol/chloroform and Centricon 100. The cDNAs were digested with Sal1 and NotI enzymes and cloned into Sal1/NotI site of pCMV-SPORT6 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by the standard chemical method. The average insert size is about 3.0kb. The library was constructed by Yulan Piao."

## ORIGIN

## Query Match

40.9%; Score 27.4; DB 7; Length 574;

Best Local Similarity 65.6%; Pred. No. 48;

Matches 40; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 5 AGCTTGAAATTAAGTCTGCCACAGTGTGTAACCTTTAGGGTTTAAGACATACAG 64

DB 303 AGTTTAAAAAAGATGATACGAGACTGAGGAAACAGTGTGTTAGGAAAGTACAG 244

QY 65 A 65

DB 243 A 243

## RESULT 15

## AA010748

## LOCUS

## DEFINITION

HS\_3067\_A1.B11.T7A.RPCL-11 Human Male BAC Library Homo sapiens genomic clone Plate=643 Col=21 Row=1, genomic survey sequence.

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## source

1. 587  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/clone="Plate=643 Col=21 Row=1"  
/sex="male"

/clone\_lib="RPCL-11 Human Male BAC Library"

/note="Vector: pBAC3.6; Site\_1: EcoRI; Site\_2: EcoRI;

Male blood DNA was isolated from one randomly chosen donor

and partially digested with a combination of EcoRI and

EcoRI methylase. Size selected DNA was cloned into the

pBAC3.6 vector at EcoRI sites"

## ORIGIN

## Query Match

40.9%; Score 27.4; DB 8; Length 587;

Best Local Similarity 69.8%; Pred. No. 49;

Matches 37; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 11 GAAATTAAGTCTGCCACAGTGTGTAACCTTTAGGGTTTAAGACATACAA 63

DB 522 GAAATTAAGTCTGCCACAGTGTGTAACCTTTAGGGTTTAAGACATACAA 574

Search completed: February 8, 2005, 22:18:16  
Job time : 806.033 secs

## FEATURES

High quality sequence atop: 587.  
Location/Qualifiers

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 8, 2005, 11:28:00 ; Search time 797.321 Seconds

(without alignments)  
8481.433 Million cell updates/sec

Title: US-10-754-437-10

Perfect score: 143  
Sequence: 1 gattctgttcgtcgtcctcc.....gtttaagacagctacaagatc 143

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_hgt:\*  
3: gb\_in:\*  
4: gb\_cm:\*  
5: gb\_ov:\*  
6: gb\_pac:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sta:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	85.4	59.7	5562	9 AB091787	AB091787 Ocolemur
2	85.4	59.7	6451	4 AB091789	AB091789 Bos tauru
3	83.8	58.6	5684	9 AB091785	AB091785 Lemur cat
4	83.4	58.3	5712	4 AB091793	AB091793 Equus cab
5	76.2	53.3	6465	9 AB091781	AB091781 Pan trogl
6	74.6	52.2	8810	9 AY040206	AY040206 Homo sapi
7	74.6	52.2	158142	2 AL357130	AL357130 Homo sapi
8	74.4	52.2	259202	9 AC002366	AC002366 Human XP2
9	74.4	52.0	685	11 BV089285	BV089285 RPRAMSECO
10	74.4	52.0	695	11 BV097603	BV097603 RPRAMSECO
11	72.8	50.9	6442	9 AB091783	AB091783 Salimiri s
12	72.2	50.5	463	10 D8306382	D83064 Mus musculi
13	72.2	50.5	9384	10 AF294397	AF294397 Mus musculi
14	72.2	50.5	95826	10 AL805974	AL805974 Mouse DNA
15	71	49.7	7425	4 AB091791	AB091791 Sus scrofa
16	68.8	48.1	177654	2 AP000918	AP000918 Homo sapi
17	68.8	48.1	200214	9 AC013412	AC013412 Homo sapi
18	68.4	47.8	212886	2 AC093946	AC093946 Rattus no
19	68.4	47.8	234471	2 AC121424	AC121424 Rattus no

c	20	68.2	47.7	363	11 BV089294	BV089294 RPRAMSECO
c	21	68.2	47.7	363	11 BV097602	BV097602 RPRAMSECO
c	22	67.2	47.0	7163	9 AB091782	AB091782 Pan trogl
c	23	67.2	47.0	38765	9 BS000568	BS000568 Pan trogl
c	24	67.2	47.0	177726	9 AC145770	AC145770 Pan trogl
c	25	67.2	47.0	190089	9 BS000576	BS000576 Pan trogl
c	26	66.4	46.4	6931	4 AB091792	AB091792 Sus scrofa
c	27	64.2	44.9	5591	4 AB091794	AB091794 Equus cab
c	28	61	42.7	8004	9 AB091786	AB091786 Lemur cat
c	29	56.8	33.7	7454	9 AB091784	AB091784 Salimiri s
c	30	47.2	33.0	6264	4 AB091790	AB091790 Bos tauru
c	31	42.6	29.8	1935	9 D83730	D83730 Homo sapien
c	32	42.4	29.7	5151	9 AB091788	AB091788 Ocolemur
c	33	41.8	29.2	727	10 S74899	S74899 amelogenin
c	34	41.8	29.2	727	10 MUSNAMEB	D31769 Mus musculi
c	35	41.8	29.2	765	4 AB032194	AB032194 Equus cab
c	36	41.8	29.2	789	4 AB032193	AB032193 Equus cab
c	37	41.8	29.2	798	10 BC059090	BC059090 Mus musculi
c	38	41.8	29.2	799	10 MUSNAMEA	D31768 Mus musculi
c	39	40.2	28.1	65	6 CQ531502	CQ531502 Sequence
c	40	40.2	28.1	359	10 RRU07054	UR07054 Rattus norv
c	41	40.2	28.1	457	10 RRU060564	UR060564 Rattus norv
c	42	40.2	28.1	476	6 AR452534	AR452534 Sequence
c	43	40.2	28.1	753	10 RRU060562	UR060562 Rattus norv
c	44	40.2	28.1	780	10 U01245	U01245 Rattus norv
c	45	40.2	28.1	812	10 RRU67130	UR67130 Rattus norv

#### ALIGNMENTS

RESULT 1  
AB091787  
LOCUS  
DEFINITION . Ocolemur garnettii AMELX gene for amelogenin, partial cds.  
ACCESSION AB091787  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Ocolemur garnettii (small-eared galago)  
Ocolemur garnettii  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Strepsithini; Galagonidae; Ocolemur.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
1 Iwase, M., Saita, Y., Hirai, Y., Hirai, H., Imai, H. and Takahata, N.  
From the cover: The amelogenin loci span an ancient pseudautosomal boundary in diverse mammalian species  
Proc. Natl. Acad. Sci. U.S.A. 100 (9), 5258-5263 (2003)  
12672962

2 (bases 1 to 5562)  
Iwase, M., Saita, Y. and Takahata, N.  
Direct Submision  
Submitted (19-SEP-2002) Mineyo Iwase, Graduate University for  
Advanced Studies (Soken-dai), Department of Biosystems Science;  
Shonan Kokuasimura, Hayama, Kanagawa 240-0193, Japan  
(E-mail: iwase@koryu.wol.soken.ac.jp, Tel: 81-468-58-1571,  
Fax: 81-468-58-1544)  
Location/Qualifiers

#### FEATURES

source  
gene  
5' UTR  
CDS  
1  
/organism="Ocolemur garnettii"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:30611"  
/sex="male"  
267..5562  
/gene="AMELX"  
join(267..321,1401..1412)  
/gene="AMELX"  
join(1413..1466,3385..3432,4699..4740,4832..4876,  
5125..5562)  
/gene="AMELX"  
/codon\_start=1  
/product="amelogenin"  
/protein\_id="BAC66107.1"







Dd		1585	CAATTTGTTTGGCTTCCTCGTGGAAGCAGCTTTGCCCATGCCCGGTGAATAAACCCTCT	1585
Oy	61	-----GCATAATATTCAATTTACAAAGCTTGGAAATATAAAGTCTGCCCAACCAAGTTGCT	114	
Dd		1586	TGCATAAGTAGCTGCATGATTTCCACAAGCTTAAGAAATATAAATCTTTCACACCGCTGGTA	1645
Oy	115	AACCTTAGCGTTTAAGACATRCAGAATC	143	
Dd		1646	AACCTTAGCGTTTAATAATGATATAATATC	1674
 RESULT 4 AB091793				
LOCUS	AB091793	5712 bp	DNA	linear
DEFINITION	Equus caballus AMELX gene for amelogenin, partial cds.	MAM 02-MAY-2003		
ACCESSION	AB091793			
VERSION	AB091793.1	GI:29126038		
KEYWORDS				
SOURCE	Equus caballus (horse)			
ORGANISM	Equus caballus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Perissodactyla; Equidae; Equus.			
REFERENCE	1 Iwase,M., Satta,Y., Hirai,Y., Hirai,H., Imai,H. and Takahata,N. From the Cover: The amelogenin loci span an ancient pseudoautosomal boundary in diverse mammalian species Proc. Natl. Acad. Sci. U.S.A. 100 (9), 5258-5263 (2003) PubMed 12672962			
JOURNAL	2 (bases 1 to 5712)			
MEDLINE	Iwase,M., Satta,Y. and Takahata,N. Direct Submission Submitted (19-SEP-2002) Mineyo Iwase, Graduate University for Advanced Studies(Sokendai), Department of Biosystems Science, Shonan Kenyakuji-mura, Hayama, Kanagawa 240-0193, Japan (E-mail:iwase@instkoryu.uoi.soken.ac.jp, Tel:81-468-58-1571, Fax:81-468-58-1544)			
REFERENCE				
AUTHORS				
TITLE				
FEATURES	Location/Qualifiers			
source	1..5712 /organism="Equus caballus" /mol_type="genomic DNA" /db_xref="taxon:9796" /sex="male" 174..5712 /gene="AMELX" join(174..233,1578..1589) /gene="AMELX" join(1580..1643,3566..3613,4993..5037,5305..>5712) /gene="AMELX" /codon_start=1 /product="amelogenin" /protein_id="BAC66112.1" /db_xref="gi:29126038" /translation="MGTWILPASLLGAAPALPLPSHPGHPEYINFSEVLTPLKWTQS LIRQPTSYGYEPWGSGHLNQIIPVLSQDNPSNHALPQHNIIPWVSQHDPVPQQPM PLPGSHSWPTQNHQPNLPVQQOPFHFQPFHQPOFPQPHQPIQOPRIHPIQLPRQPLRP PIPLQRLRPLRLPLRLAMPATEQ"			
gene				
5'UTR				
CDS				
ORIGIN				
Query Match	58.3%	Score 83.4	DB 4,	Length 5712;
Best Local Similarity	79.7%	Pred. No. 1.8e-15;		
Matches 114;	Conservative 0;	No.matches 21;	Indels .8;	Gaps 1;
Oy	1	GATTTGTTGGCGGCTCTCGGAGACGCGCTTGCTATGCCCGCTGATAAAAATACCCCT	60	
Dd	1601	GATTTGTTGGCGACGCTCTCGGAGACGCGCTTGCTATAACCCGTGATATGATC----	1656	
Oy	61	GCATATATTCATTTCAAGCTTGGAAATATAAAGTCTGCCCAACCAAGTTGTAATCTTT	120	
Dd	1657	-----AAATCCATTTTCCAAAGCTTTGAAAAATACAAATCTGCTCCACGGTTGTAATCTTT	1712	
Oy	121	AGGGTTTAAGACATGACAAGATC	143	

[illegible]

```

DEFINITION Homo sapiens amelogenin precursor (AMELX) gene, complete cds.
ACCESSION AY040206
VERSION AY040206.1 GI:15028582
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Hart, P.S., Hart, T.C., Simmer, J.P. and Wright, J.T.
TITLE A nomenclature for X-linked amelogensis imperfecta
JOURNAL Arch. Oral Biol. 47 (4), 255-260 (2002)
MEDLINE 21920287
PUBMED 11922868
REFERENCE
AUTHORS Hart, S., Hart, T.C. and Wright, T.
TITLE Direct Submission
JOURNAL Submitted (14-JUN-2001) Human Genetics, University of Pittsburgh,
3550 Terrace St., 572A Scaife Hall, Pittsburgh, PA 15090, USA
FEATURES
SOURCE
1..8810
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="X"
/map="Xp22.3-p22.1"
1..8791
/gene="AMELX"
1..1416
/gene="AMELX"
1386..1391
/gene="AMELX"
join(1417..1472,2781..2846,4783..4830,6240..6281,6552,6977,8611..8791)
/gene="AMELX"
/product="amelogenin precursor"
1417..1472
/gene="AMELX"
/number=1
2781..2846
/gene="AMELX"
/number=2
join(2793..2846,4783..4830,6240..6281,6552,6977,8611..8616)
/gene="AMELX"
/codon_start=1
/product="amelogenin precursor"
/db_xref="GI:15028583"
/translation="MGTWLIFACILGAAPAPLPPHGHGYINFSEVLTPLKWTOS
IRPPYSYGYEPMGWLHIOIIVLSQHPPTHTLQPHHPIVVPACQOVIPQPPMP
VPCGSHMTPIOHQPNLPPACQOVPOPOVQPPHQPMPQPPVHMQPLPPQPLPP
MPMPQPLPMLPLTLEAWPSTDKRREVD"
2793..2840
/gene="AMELX"
4783..4830
/gene="AMELX"
/number=3
6107..6148
/gene="AMELX"
/number=4
6240..6281
/gene="AMELX"
/number=5
6552..6977
/gene="AMELX"
/number=6
8611..8791
/gene="AMELX"
/number=7
8785..8791
/gene="AMELX"
sig_peptide
exon
exon
exon
exon
exon
exon
polya_signal
ORIGIN

```

```

Query Match 52.2%; Score 74.6; DB 9; Length 8810;
Best Local Similarity 78.5%; Pred. No. 1.1e-12;
Matches 117; Conservative 0; Mismatches 24; Indels 8; Gaps 2;
QY 1 GATTTGTTGGCTGCTTCCTGAGACACCTTTGCTAGCCCGAGTAAATACCCCT 60
DB 2804 GATTTATTTGCTGCTTCCTGAGACACCTTTGCTAGCCCGAGTAAATACCCCT 2863
QY 61 -----GCATTAATATTCATTTACAGCTTGAATAAAGCTGCCACCACTTGGT 114
DB 2864 TGCAATAGCATGTCATTTACAACTTGACATTAATCTGCTC--ATAGTTGCT 2921
QY 115 AACTTAGCGTTTAAACAGTACAGATC 143
DB 2922 GAAATTAGCGTTTAAACAGTATGAGATC 2950
RESULT 7
AL357130/c AL357130 158142 bp DNA linear HTG 10-JUL-2001
LOCUS Homo sapiens chromosome X clone RP13-169E15, 4 unordered pieces.
DEFINITION AL357130
ACCESSION AL357130.3 GI:9863814
VERSION HTG: HTGS_PHASE1: HTGS_CANCELLED.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS McIay, K.
TITLE Direct Submission
JOURNAL Submitted (09-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
COMMENT On Aug 21, 2000 this sequence version replaced gi:9214076.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: b169E15
----- Summary Statistics
Assembly program: XGAP4; Version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 155448 bases at least Q40
Consensus quality: 156466 bases at least Q30
Insert size: 157842; sum-of-contigs
Insert size: 160705; 33.4% error; agarose-fp
Quality coverage: 4.56x in Q20 bases; sum-of-contigs Quality
coverage: 4.56x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 68897: contig of 68897 bp in length
* 68898 68997: gap of 100 bp
* 68998 122842: contig of 53845 bp in length
* 122843 122942: gap of 100 bp
* 122943 123584: contig of 6642 bp in length
* 123585 123685: gap of 100 bp
* 123685 158142: contig of 28458 bp in length.
Location/Qualifiers
1..158142
/organism="Homo sapiens"

```

```

/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="X"
/clone="RP13-169E15"
/clone_id="RP13-13.1"
1. 688937
/misc_feature
/note="assembly_fragment:01108
fragment_chain:1
clone_end:r7
vector_side:left"
68998..122842
/misc_feature
/note="assembly_fragment:00101
fragment_chain:1"
122943..129584
/misc_feature
/note="assembly_fragment:00361
fragment_chain:1"
129685..158142
/misc_feature
/note="assembly_fragment:01285
fragment_chain:1
clone_end:SP6
vector_side:right"

```

Query Match	52.2%	Score 74.6	DB 2	Length 158142
Best Local Similarity	78.5%	Pred. No. 1.1e-12		
Matches 117	Conservative 0	Mismatches 24	Indels 8	Gaps 2
Qy	1	GATTTTGTTCCTGCCTCCTGGAGCAGCCTTGTATGCCGTGAGTAAATACCCCT	60	
Db	44314	GATTTTATTTCTGCTCCTCGGAGCAGCTTTTGCATGCTGTGATGTAACACCCCT	44255	
Qy	61	-----GCATAATATTCATTTCACAAGCTTGAAATAAAGTTGCCCCACCACTTGGT	114	
Db	44254	TGCATTAAGTCAGTGTCCAAATTCACAACTTGACATMAAATCTGCT--ATATGGGT	44197	
Qy	115	AACCTTAGGGTTTAAAGACAGTACAAGATC	143	
Db	44196	GAAATTAGGGTTTAAACAGATGATGATC	44168	
RESULT 8	AC002366	259202 bp	DNA	linear
LOCUS	AC002366	259202 bp	DNA	linear
DEFINITION	Human Xp22 BAC CT-285115 (from Caltech/Research Genetics), PAC			
	RPC11-27C2 (from Roswell Park Cancer Center), and Cosmid U35B5			
	(from Lawrence Livermore), complete sequence.			
ACCESSION	AC002366	U79549	U70036	
VERSION	AC002366.1	GI:2739349		
KEYWORDS	HTG.			
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			
	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
REFERENCE	1 (bases 1 to 259202)			
AUTHORS	Murny,D., Anand-Lari,M.A., Timms,K.M., Yu,W., Dugan,S., Lu,J.,			
	Shen,Y., Rowland,K., Liu,W., Perez,L., Ding,Y., Haywood,M.,			
	Ustin,A., Leal,B., Logan,O., Nguyen,V., Savage,L., Shen,H.,			
	Worley,K., Chen,E., Forcun,J., Aronson,A.D., Chiu,M.W.,			
	Correll,J.H., Brundage,E., Di,W., Chinnault,C., Nelson,D. and			
	Gibbs,R.A.			
	Direct Submission			
TITLE	Unpublished			
JOURNAL	2 (bases 1 to 259202)			
REFERENCE	Chiu,M.W.			
AUTHORS	Direct Submission			
TITLE	Submitted (23-JUL-1997)			
JOURNAL	Molecular and Human Genetics, Baylor			
	College of Medicine, One Baylor Plaza, Houston, TX 77030, USA			
REFERENCE	3 (bases 1 to 259202)			
AUTHORS	Chiu,M.W.			
TITLE	Direct Submission			
JOURNAL	Submitted (02-JAN-1998)			
	Molecular and Human Genetics, Baylor			
	College of Medicine, One Baylor Plaza, Houston, TX 77030, USA			
REFERENCE	4 (bases 1 to 259202)			

AUTHORS	TITLE	JOURNAL	COMMENT
Chiu,M.W.	Direct Submission	Submitted (11-JUN-1998)	Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
		On Jan 2, 1998	this sequence version replaced gi:2642176.
			Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality does not meet this standard, it will be indicated in the annotation.
			The repeat regions shown were identified using RepeatMasker by Adrian Smlt.
			Sequence similarities were identified using Powerblast by Jinghui Zhang.
			Exon/intron boundaries of identified genes were chosen if there were canonical splice junctions that maintained sequence continuity across the splice junctions.
FEATURES			Location/Qualifiers
source			1. .259202
			/organism="Homo sapiens"
			/mol_type="genomic DNA"
			/db_xref="taxon:9606"
			/chromosome="X"
			/map="Xp22"
			/clone="CT-285115, RPTC11-27C22, U35B5"
			/complement(1. .147)
			/note="overlap 1. .147 in AC004467"
misc_feature			35. .127
repeat_region			/rpt_family="MIR"
			/complement(1087. .1502)
repeat_region			/rpt_family="L1"
			/complement(5292. .5579)
repeat_region			/rpt_family="Alu"
			5302. .5553
repeat_region			/rpt_family="SVA"
			/complement(6009. .6063)
repeat_region			/rpt_family="L1MC2"
			/complement(7092. .7382)
repeat_region			/rpt_family="L1M55"
			7423. .7721
repeat_region			/rpt_family="Alu"
			9771. .10072
repeat_region			/rpt_family="L1MC2"
			/complement(10072. .11146)
repeat_region			/rpt_family="THP"
			/complement(11156. .11528)
repeat_region			/rpt_family="THE1b"
			/complement(12944. .13151)
repeat_region			/rpt_family="MER20"
			13862. .14010
repeat_region			/rpt_family="MLT1f"
			14063. .14352
repeat_region			/rpt_family="MER42a"
			15847. .16170
repeat_region			/rpt_family="Alu"
			/complement(18743. .19117)
repeat_region			/rpt_family="MER7"
			23507. .23614
repeat_region			/rpt_family="L1"
			23615. .24425
repeat_region			/rpt_family="L1PA11"
			/complement(25475. .25968)
repeat_region			/rpt_family="MLT1d"
			27234. .27683
STS			/note="similar to EST with GenBank Accession Number H48827"
repeat_region			/db_xref="dbSTS:G26372"
			/complement(28186. .28366)
repeat_region			/rpt_family="MER20"
			28715. .28995

```

repeat_region      /rpt_family="Alu"
                   complement(30217, .30408)
repeat_region      /rpt_family="L1M4S"
                   31195, .31364
repeat_region      /rpt_family="Alu"
                   complement(33939, .34460)
repeat_region      /rpt_family="L1M42"
                   34952, .35384
repeat_region      /rpt_family="MLT1d"
                   36035, .36126
repeat_region      /rpt_family="MIR2"
                   36850, .37035
repeat_region      /rpt_family="MER20"
                   38042, .38250
STS               /db_xref="dbSTS:L47508"
repeat_region      38122, .38510
repeat_region      /rpt_family="MIRa"
                   38797, .40103
repeat_region      /rpt_family="MIR"
                   complement(42465, .42539)
repeat_region      /rpt_family="MIR"
                   47139, .47335
repeat_region      /rpt_family="MER3"
                   complement(48993, .49166)
repeat_region      /rpt_family="MIR"
                   complement(49718, .49918)
repeat_region      /rpt_family="L1P411"
                   complement(51943, .52505)
repeat_region      /rpt_family="L1MB3"
                   53226, .53677
repeat_region      /rpt_family="MLT1d"
                   complement(54846, .54966)
repeat_region      /rpt_family="MIR"
                   55040, .55226
repeat_region      /rpt_family="MER20"
                   55691, .55837
repeat_region      /rpt_family="Alu"
                   56051, .56486
repeat_region      /rpt_family="L1MB3a"
                   58557, .59030
repeat_region      /rpt_family="MLT1d"
                   complement(59270, .59428)
repeat_region      /rpt_family="MER5"
                   complement(59847, .60030)
repeat_region      /rpt_family="MER20"
                   complement(61290, .61384)
repeat_region      /rpt_family="L1PB1"
                   61683, .61811
repeat_region      /rpt_family="MER20"
                   61907, .62113
repeat_region      /rpt_family="L1MB7"
                   complement(63588, .63969)
repeat_region      /rpt_family="L1MA10"
                   64800, .65073
repeat_region      /rpt_family="Alu"
                   complement(67157, .67209)
repeat_region      /rpt_family="MIR"
                   67637, .75845
Gene              /gene="HAMEIX"
Promoter          67637, .68496
                  /gene="HAMEIX"
MRNA              /note="similar to bovine amelogenin promoter"
                  join(68495, .68552,69861, .69926,71863, .71910,73320, .73361,
73632, .74057,75691, .75845)
                  /gene="HAMEIX"
                  /product="amelogenin"
                  join(69473, .69926,71863, .71910,73320, .73361,74057,
75691, .75696)
                  /gene="HAMEIX"
                  /note="similar to amelogenin encoded by GenBank Accession
Number M86932"
                  /codon_start=1
                  /product="amelogenin"

```

```

/protein_id="AAC21581.1"
/db_xref="gi:2739350"
/translation="MGTWIIIPACILGAFAMLPLRYHNGHCXILSFYEVLTPKLYQAS
IRPRSGVSGYPRMGKYLHNOILPYLSOHNPRFLODNNHILPVVRAQDQPIVPODPMR
VGGHSTHPIEDHQNPLRRPAQDPIQFORQVORHNOPLRVIRVORMLRORPLRR
MEVQPLRPMLRDLTILAMBSTDTKKEVD"
repeat_region
  complement(77161..77331)
  /rpt_family="MER3"
  complement(79184..79473)
  /rpt_family="Alu"
repeat_region
  complement(79881..80169)
  /rpt_family="Alu"
repeat_region
  80156..80253
  /rpt_family="MER20"
repeat_region
  80384..80459
  /rpt_family="L1PB3"
repeat_region
  complement(81544..81633)
  /rpt_family="MIR"
repeat_region
  84789..84957
  /rpt_family="L1MA2"
repeat_region
  84978..85352
  /rpt_family="Alu"
repeat_region
  85294..85692

```

	Query March Best Local Similarity Matches 117; Conservative	52.2%; 78.5%; 0;	Score 74.6; Pred. No. 1,1e-12; Mismatches 24;	DB 9; Length 259202; Indels 8; Gaps 2;
QY	1 GATTTTGTTCCTGCCTCCTGGAGACAGCTTGTCTATGCCCGTAGTAATAATACCCCT			60
Db	69884 GATTTTATTTTGCTGCCTCCTGGAGACAGCTTTTGCATGCGCTGTGATTAACAACCCCT			69943
QY	61 -----GCATTAATATTCATTTTCAACAAGCTTGGAAATTAAGTCTGCCCAACAGTTGCT			114
Db	69944 TGCATAGTCAGTCAGTCCAAATTCACAACTTGGACATATAAAATCTGCTC--ATAGTTGCT			70001
QY	115 AACCTTAGGGTTTAAGACAGTACAAATC			143
Db	70002 GAAATTTAGGGTTTAAACAGTATGAAATC			70030

RESULT 9	
LOCUS	BV089295/c
DEFINITION	695 bp DNA linear STS 15-OCT-2003
ACCESSION	U00000001191 Roche Palo Alto Mus musculus STS genomic, sequence
VERSION	U00000001191 Roche Palo Alto Mus musculus STS genomic, sequence
KEYWORDS	tagged site.
SOURCE	BV089295.1 GI:37666774
ORGANISM	Mus musculus (house mouse)
REFERENCE	Mus musculus
AUTHORS	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 695)
TITLE	Uauke, J., Liao, G., Cheng, J., Nguyen, A., Bach, C., Puech, A.,
JOURNAL	McPherson, J. D., Foernzier, D. and Peltz, G.
COMMENT	Mus musculus SNPs
	Unpublished (2003)

CONTACT: Jonathan Usuka  
Roche Palo Alto Genetics and Genomics Department  
Roche Palo Alto  
3431 Hillview Ave, Mailstop S3-1, Palo Alto, CA 94024, USA  
Tel: 6508855807  
Email: Jonathan.Usuka@roche.com  
Primer A: No primer submitted  
Primer B: No primer submitted.  
Location/Qualifiers

```

source
1. .695
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/map="X-73613-72919-AL085974.7.1.95826"

```

/clone lib="Roche Palo Alto"  
 /note="SNPs developed from assay sequences derived from 15 different strains of mice (as of October 1, 2003). Those strains include A/J, A/HeJ, -129/Sv, AKR/J, B10.D2-H2/oshu, BALB/cByJ, BALB/cJ, C3H/HeJ, C57BL/6J, -CAST/Ei, DBA/2J, MRJ/MPJ, NZB/BlmJ, NZW/Lac, SPRET/Ei." <1..>695

## ORIGIN

Query Match 52.0%; Score 74.4; DB 11; Length 695;  
 Best Local Similarity 77.9%; Pred. No. 1.2e-12;  
 Matches 116; Conservative 0; Mismatches 27; Indels 6; Gaps 2;

QY 1 GATTGGTTCCTGCTGCTGGAGAGCCTTGTGATCCCGTAGTAATAATACCC- 59  
 DB 479 GATTGGTTCCTGCTGCTGGAGAGCCTTGTGATCCCGTAGTAATAATACCA 420  
 QY 60 ----TGCATATATTCATTTCAACAAGCTTGAATAAAGTGTGCCCCACAGTTGTA 115  
 DB 419 TTACTAATTCATTCATTCATTAATCTTGAATGAAATCTGCTCCCGTAGTGTGA 360  
 QY 116 ACTTT-AGGCTTTAAGACAGTACAGATC 143  
 DB 359 ATTTCACGTGTGTACACAGTCAAGATC 331

RESULT 10 695 bp DNA linear STS 14-JAN-2004  
 BV097603 Roche Palo Alto Mus musculus STS genomic, sequence  
 LOCUS BV097603  
 DEFINITION tagged site.

ACCESSION BV097603  
 VERSION BV097603.1 GI:40810679  
 KEYWORDS STS.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 695)  
 Uenka, J., Iiao, G., Cheng, J., Nguyen, A., Bach, C., Puech, A., McPherson, J.D., Foerzler, D. and Peltz, G.  
 Mus musculus SNPs  
 Unpublished (2003)

## COMMENT

Contact: Jonathan Uenka  
 Roche Palo Alto Genetics and Genomics Department  
 Roche Palo Alto  
 3431 Hillview Ave, Mailstop S3-1, Palo Alto, CA 94024, USA  
 Tel: 6508555807  
 Email: jonathan.uenka@roche.com  
 Primer A: No primer submitted with this STS  
 Primer B: No primer submitted with this STS.  
 Location/Qualifiers

## FEATURES

source  
 1..695  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10090"  
 /map="X-73613-72919-AL805974.7.1.95826"  
 /clone.lib="Roche Palo Alto"  
 /note="SNPs developed from assay sequences derived from 15 different strains of mice (as of October 1, 2003). Those strains include A/J, A/HeJ, 129/Sv, AKR/J, B10.D2-H2/oshu, BALB/cByJ, BALB/cJ, C3H/HeJ, C57BL/6J, CAST/Ei, DBA/2J, MRJ/MPJ, NZB/BlmJ, NZW/Lac, SPRET/Ei." <1..>695

## ORIGIN

Query Match 52.0%; Score 74.4; DB 11; Length 695;  
 Best Local Similarity 77.9%; Pred. No. 1.2e-12;  
 Matches 116; Conservative 0; Mismatches 27; Indels 6; Gaps 2;  
 QY 1 GATTGGTTCCTGCTGCTGGAGAGCCTTGTGATCCCGTAGTAATAATACCC- 59  
 DB 479 GATTGGTTCCTGCTGCTGGAGAGCCTTGTGATCCCGTAGTAATAATACCA 420

DB 479 GATTGGTTCCTGCTGCTGGAGAGCCTTGTGATCCCGTAGTAATAATACCA 420  
 QY 60 ----TGCATATATTCATTTCAACAAGCTTGAATAAAGTGTGCCCCACAGTTGTA 115  
 DB 419 TTACTAATTCATTCATTCATTAATCTTGAATGAAATCTGCTCCCGTAGTGTGA 360  
 QY 116 ACTTT-AGGCTTTAAGACAGTACAGATC 143  
 DB 359 ATTTCACGTGTGTACACAGTCAAGATC 331

RESULT 11 6442 bp DNA linear PRI 02-MAY-2003  
 AB091783  
 LOCUS AB091783  
 DEFINITION Saimiri sclureus AMELX gene for amelogenin, partial cde.  
 ACCESSION AB091783.1 GI:29126019  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

Saimiri sclureus (common squirrel monkey)  
 Saimiri sclureus  
 Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Cebinae;  
 Saimiri.

REFERENCE Iwase, M., Satta, Y., Hirai, Y., Hirai, H., Imai, H. and Takahata, N.  
 TITLE From the Cover: The amelogenin loci span an ancient pseudautosomal boundary in diverse mammalian species  
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 100 (9), 5258-5263 (2003)  
 MEDLINE 22608569  
 PUBMED 12672962

2 (bases 1 to 6442)  
 Iwase, M., Satta, Y. and Takahata, N.  
 Direct Submission  
 Submitted (19-SEP-2002) Mineyo Iwase, Graduate University for Advanced Studies (Sokenai), Department of Biosystems Science; Shonan Kokuai-mura, Hayama, Kanagawa 240-0193, Japan  
 (E-mail: iwase@minyo.koryuwo1.soken.ac.jp, Tel: 81-468-58-1571, Fax: 81-468-58-1544)

Location/Qualifiers  
 1..6442  
 /organism="Saimiri sclureus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9521"  
 /sex="male"  
 /gene="AMELX"  
 /join(936..991,2298..2309)  
 /gene="AMELX"  
 /join(2310..2363,4289..4336,5592..5633,5725..5769,6041..6442)  
 /gene="AMELX"  
 /codon\_start=1  
 /product="amelogenin"  
 /protein\_id="BA066103.1"  
 /db\_xref="GI:29126020"  
 /translation="MGTWILFAAGLGAATAMPPLPHPGHYINFSYENSHQALNIDRTGLVTLPLKWTQSMIRAPVPSGYEDPKGMGLHIOIIPVLSQOHPTHTLOPHNIPVVAQOQVVPQOQPMMPVPGQSHMIPGQHOPNPPRPAQOPFOQPTVQPOHNPQOPVHMQPLPQPPLPMPFMPQPLPPLPMLPDLPLKMPRT"

## FEATURES

source  
 1..6442  
 /organism="Saimiri sclureus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9521"  
 /sex="male"  
 /gene="AMELX"  
 /join(936..991,2298..2309)  
 /gene="AMELX"  
 /join(2310..2363,4289..4336,5592..5633,5725..5769,6041..6442)  
 /gene="AMELX"  
 /codon\_start=1  
 /product="amelogenin"  
 /protein\_id="BA066103.1"  
 /db\_xref="GI:29126020"  
 /translation="MGTWILFAAGLGAATAMPPLPHPGHYINFSYENSHQALNIDRTGLVTLPLKWTQSMIRAPVPSGYEDPKGMGLHIOIIPVLSQOHPTHTLOPHNIPVVAQOQVVPQOQPMMPVPGQSHMIPGQHOPNPPRPAQOPFOQPTVQPOHNPQOPVHMQPLPQPPLPMPFMPQPLPPLPMLPDLPLKMPRT"

## ORIGIN

Query Match 50.9%; Score 72.8; DB 9; Length 6442;  
 Best Local Similarity 78.1%; Pred. No. 4e-12;  
 Matches 114; Conservative 0; Mismatches 27; Indels 5; Gaps 2;  
 QY 1 GATTGGTTCCTGCTGCTGGAGAGCCTTGTGATCCCGTAGTAATAATACCC- 60  
 DB 2321 GATTGGTTCCTGCTGCTGGAGAGCCTTGTGATCCCGTAGTAATAATACCC- 2380  
 QY 61 GCATATTA----TTCAATTTCAACAAGCTTGAATAAAGTGTGCCCCACAGTTGTA 116  
 DB 2381 CATTACTAGTGTGCAATTTCAATTAATTAAGCAATTAATTAAGTGTGCCCCA-TGGTTGTGA 2439

OY 117 CTTAGGTTTAAGACAGTACAGAT 142  
 |||||  
 Db 2440 AGTTAGGTTTAAACGTTATGAGAT 2465

RESULT 12  
 D83063S2 463 bp DNA linear ROD 26-NOV-2003  
 LOCUS Mus musculus DNA for amelogenin, exon 2.  
 DEFINITION D83064  
 ACCESSION D83064.1 GI:2687869  
 VERSION  
 KEYWORDS  
 SEGMENT  
 SOURCE  
 ORGANISM  
 Mus musculus (house mouse)  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
 AUTHORS Oida,S., Miyazaki,H., Iimura,T., Suzuki,M., Sasaki,S. and Shimokawa,H.  
 TITLE Molecular structure of the mouse amelogenin genomic DNA  
 JOURNAL DNA Seq. 6 (5), 307-310 (1996)  
 MEDLINE 97142134  
 PUBMED 8988368  
 REFERENCE 2 (bases 1 to 463)  
 AUTHORS Oida,S.  
 TITLE Direct Submision  
 JOURNAL Submitted (14-JAN-1996) Shinichiro Oida, Tokyo Medical and Dental University, Fac. Dentistry, Dept. Biochemistry, 1-5-45 Yushima, Bunkyo-ku, Tokyo 113, Japan (E-mail:s.oida.obchdent.tmd.ac.jp, Tel:03-5803-5448, Fax:03-5803-0187)

COMMENT On Dec 15, 1997 this sequence version replaced gi:2662357.  
 FEATURES  
 source  
 Location/Qualifiers  
 1..463  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /sub\_species="Balb-C"  
 /db\_xref="taxon:10090"  
 /chromosome="X"  
 /map="Xp3-F4"  
 /clone="MAMG-2"  
 /cell\_type="leukocyte"  
 /clone\_lib="EMBL-3 SP6-T7"  
 /dev\_stage="adult"  
 join(D83063.1:203..258,222..233)  
 <1..221  
 /number=1  
 222..287  
 /product="amelogenin"  
 /number=2  
 288..463  
 /number=2

ORIGIN  
 Intron  
 5'UTR  
 Intron  
 exon  
 Intron

Query Match 50.5%; Score 72.2; DB 10; Length 463;  
 Best Local Similarity 77.2%; Pred. No. 6,1e-12;  
 Matches 115; Conservative 0; Mismatches 28; Indels 6; Gaps 2;

OY 1 GATTTGTTGCTGCTGCTGAGAGAGCTTTGCTATGCCGTGATTAATACCC- 59  
 |||||  
 Db 245 GATTTGTTGCTGCTGCTGAGAGAGCTTTGCTATGCCGTGATTAATACCA 304  
 |||||  
 OY 60 ----TGCATTAATTTCAATTTCACAGCTTGAATAAAGTCTGCCCAACGATTGTA 115  
 |||||  
 Db 305 TTACTAATTTGCATTCATTAATTAATCTTGAATGAATCTGCTCCATAGTTGTA 364  
 |||||  
 OY 116 ACTTT-AGGTTTAAGACAGTACAGATC 143  
 |||||  
 Db 365 ATTTCACTGTGTAAACAGTGCAGATC 393  
 |||||

RESULT 13  
 AF294397

LOCUS AF294397 9384 bp DNA linear ROD 03-SEP-2000  
 DEFINITION Mus musculus amelogenin gene, promoter and partial cds.  
 ACCESSION AF294397  
 VERSION AF294397.1 GI:9965405  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Mus musculus (house mouse)  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
 AUTHORS Sneed,M.L., Paine,M.L., Luo,W., Zhu,D.H., Yoshida,B., Lei,Y.P., Paine,C.T., Chen,L.S., Burslein,J.M., Jitpukdeeputintara,S., White,S.N. and Bringsas,P. Jr.  
 TITLE Transgene animal model for protein expression and accumulation into forming enamel  
 JOURNAL Connect. Tissue Res. 38 (1-4), 279-286 (1998)  
 MEDLINE 20515040  
 PUBMED 11063035

REFERENCE 2 (bases 1 to 9384)  
 AUTHORS Sneed,M.L., Zhu,D.-H., Lei,Y.-P. and Paine,M.L.  
 TITLE Direct Submision  
 JOURNAL Submitted (08-AUG-2000) Dentistry, University of Southern California, 2250 Alcazar Street, CSA103, Los Angeles, CA 90033, USA  
 Location/Qualifiers  
 1..9384  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10090"  
 /chromosome="X"  
 1..8546  
 7188..7194  
 8547..8612  
 /product="amelogenin"  
 8559..8612  
 /note="enamel matrix protein"  
 /codon\_start=1  
 /product="amelogenin"  
 /protein\_id="AAG10078.1"  
 /db\_xref="GI:9965405"  
 /translation="MGWILPACLLGAAMP"

Promoter  
 TATA\_signal  
 mRNA

ORIGIN

Query Match 50.5%; Score 72.2; DB 10; Length 9384;  
 Best Local Similarity 77.2%; Pred. No. 6,3e-12;  
 Matches 115; Conservative 0; Mismatches 28; Indels 6; Gaps 2;

OY 1 GATTTGTTGCTGCTGCTGAGAGAGCTTTGCTATGCCGTGATTAATACCC- 59  
 |||||  
 Db 8570 GATTTGTTGCTGCTGCTGAGAGAGCTTTGCTATGCCGTGATTAATACCA 8629  
 |||||  
 OY 60 ----TGCATTAATTTCAATTTCACAGCTTGAATAAAGTCTGCCCAACGATTGTA 115  
 |||||  
 Db 8630 TTACTAATTTGCATTCATTAATTAATCTTGAATGAATCTGCTCCATAGTTGTA 8689  
 |||||  
 OY 116 ACTTT-AGGTTTAAGACAGTACAGATC 143  
 |||||  
 Db 8690 ATTTCACTGTGTAAACAGTGCAGATC 8718  
 |||||

RESULT 14  
 AL805974 95826 bp DNA linear ROD 29-AUG-2003  
 LOCUS AL805974/c  
 DEFINITION Mouse DNA sequence from clone RP23-334F21 on chromosome X, complete sequence.  
 ACCESSION AL805974 GI:34366495  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Mus musculus (house mouse)  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
 AUTHORS Chapman,J.



**This Page Blank (uspto)**







PT	connective tissue and cell proliferative disorders including cancer.
PS	Claim 5; SEQ ID NO 55; 272pp; English.
XX	
CC	The present sequence encodes a human cell adhesion and extracellular
CC	matrix protein designated CADECM. CADECM sequences has neuroprotective,
CC	cystostatic and anorectic activites. The CADECM polypeptides and
CC	polynucleotides are useful in diagnosing, treating and preventing immune
CC	neurological, developmental, connective tissue and cell proliferative
CC	disorders including cancer, e.g. breast, prostate, ovarian, lung or colorectal
CC	cancer, obesity and tangle disease.
XX	
SQ	Sequence 549 BP; 134 A; 203 C; 112 G; 100 T; 0 U; 0 Other;
Query Match	28.3%; Score 40.4; DB 12; Length 549;
Best Local Similarity	88.0%; Pred. No. 0.0012;
Matches 44; Conservative	0; Mismatches 6; Indels 0; Gaps 0
Oy	1 GATTTCGCTTGCCTCCTCGGACAGCCTTGCAATGCCGTAGTA 50       47 GATTTCGCTTGCCTCCTCGGACAGCCTTGCAATGCCGTAGTA 96
Dd	
RESULT 4	
ADM80827	
ID	ADM80827 standard; cDNA; 623 BP.
XX	
AC	ADM80827;
XX	
DT	03-JUN-2004 (first entry)
XX	
DE	Human CADECM-14 encoding cDNA SEQ ID NO:56.
XX	
KW	human; cell adhesion and extracellular matrix protein; CADECM;
KV	neuroprotective; cystostatic; anorectic; immune disorder;
KM	neurological disorder; developmental disorder;
KX	connective tissue disorder; cell proliferative disorder; cancer; obesity;
XX	tangle disease; gene; ss.
XX	
OS	Homo sapiens.
XX	
PH	
FT	Key Location/Qualifiers
CDS	36..563
FT	/tag= a
FT	/product= "CADECM-14"
XX	
PM	M02004015396-A2.
XX	
PD	19-FEB-2004.
XX	
PF	12-AUG-2003; 2003WO-US025418.
XX	
PR	13-AUG-2002; 2002US-0403781P.
PR	30-AUG-2002; 2002US-0407034P.
PR	13-SEP-2002; 2002US-0410566P.
PR	24-SEP-2002; 2002US-0413482P.
PR	05-SEP-2002; 2002US-0413890P.
PR	08-NOV-2002; 2002US-0424904P.
PR	13-NOV-2002; 2002US-0426222P.
XX	
PA	(INCY-) INCYTE CORP.
XX	
P1	Elliott VS, Khare R, Emerling BM, Kable AB, Tran UK, Jin P;
P1	Beccha SD, Marcquis JP, Swarnakar A, Chewla NK, Ramkumar J;
P1	Hafalla AD, Lee SY, Jiang X, Jackson AA, Richardson TW, Blake JJ;
P1	Wang JT, Chien D, Yang YG;
XX	
DR	WPI; 2004-191795/18.
XX	
DR	P-PsDB; ADM80785.
XX	
PT	New cell adhesion and extracellular matrix proteins, useful in
PT	diagnosing, treating and preventing immune, neurological, developmental,
PT	connective tissue and cell proliferative disorders including cancer.

XX Claim 5; SEQ ID NO 56; 272bp; English.

PS

CC The present sequence encodes a human cell adhesion and extracellular  
CC matrix protein designated CADECM. CADECM sequences has neuroprotective,  
CC cytoskeletal and anorectic activities. The CADECM polypeptides and  
CC polynucleotides are useful in diagnosing, treating and preventing immune,  
CC neurological, developmental, connective tissue and cell proliferative  
CC disorders including cancer, e.g. breast, prostate, ovarian, lung or colon  
CC cancer, obesity and Tangier disease.  
CX

SQ Sequence 623 BP; 150 A; 228 C; 128 G; 117 T; 0 U; 0 Other;

Query Match      28.3%; Score 40.4; DB 12; Length 623;  
Best Local Similarity    88.0%; Pred. No. 0.0013;  
Matches    44; Conservative    0; Mismatches    6; Indels    0; Gaps    0;

Gy                 1 GATTGTGTTGCCGCCTCGGGAGACGCTTTCATACCCTAAGTA 50  
DB                 47 GATTTATATGCGCTCCTCGAGAGACTTTTGCAATCGCTGCTTA 96

RESULT 5  
ABN28389  
ID ABN28389 standard; DNA; 65 BP.  
XX  
XX AC ABN28389;  
XX  
XX DT 15-JUL-2002 (first entry)  
XX  
XX DE Rat spliced transcript detection oligonucleotide SEQ ID NO:1137.  
XX  
XX Human; mouse; rat; splice transcript; detection; RNA transcript;  
KW splunc variant; transcripome; oligonucleotide library; ss.  
KM  
XX Rattus norvegicus.  
OS  
  
PM WO200210449-A2.  
PN  
PD 07-FEB-2002.  
PP  
XX 20-JUL-2001; 2001WO-IB001903.  
PR  
PR 28-JUN-2000; 2000US-0221607P.  
PR 02-MAY-2001; 2001US-0287724P.  
XX  
PA (COMP-) COMPUGEN INC.  
XX  
PI Shoshan A, Wasserman A, Mintz E, Mintz L, Faigler S;  
PI WPI; 2002-257383/30.  
DR  
XX  
XX New oligonucleotide libraries comprising oligonucleotides which  
PT selectively hybridize to mRNAs transcribed from a transcription unit of a  
PT genome, useful for detecting tissue-, pathology-, and developmental-specific genes.  
PX  
XX Example 1; SEQ ID NO 1137; 47bp; English.

The present invention describes oligonucleotide libraries for detecting messenger RNAs that populate a (sub-)transcriptome, where the (sub-)transcriptome comprises messenger RNAs transcribed from multiple transcription units that populate a genome. The library comprises several oligonucleotides, each capable of hybridizing selectively to a set of messenger RNAs transcribed from a given transcription unit of the genome, which encodes one or more messenger RNA splice variants. The oligonucleotide libraries are useful for detecting mRNAs from a biological sample, in expression profiling studies, in qualitatively or quantitatively characterizing the corresponding transcriptome, and in detecting RNA transcripts and splice variants of human or animal transcriptomes. The libraries may also be used as specialised mini-libraries to detect transcripts of a sub-transcriptome under a particular biological or pathological state, and so allowing the detection of tissue

Query Match	28.1%	Score 40.2	DB 6	Length 65
Best Local Similarity	93.3%	Pred. No. 0.00067		
Matches 42	Conservative 0	Mismatches 3	Indels 0	Gaps 0
1	GATTTGTTTCCCTGCGCCCTCGGAGCAGCCTTTGCTATGCCCGT	45		
6	GATCTTGTTGCTGCTCCCTCGGAGCAGCCTTTGCTATGCCCGT	50		
<p>RESULT 6 AAZ50832 AAZ50832 standard; DNA; 476 BP.</p> <p>31-MAY-2000 (first entry)</p> <p>Rat amelogenin gene (AA).</p> <p>Amelogenin; splice variant; rat; (AA); chondrogenesis; osteogenesis; chondrogenic inducing molecule; CIM; cartilage growth; osteopathic; extracellular matrix protein; tooth enamel; enamel mineralisation; ameloblast; bone regeneration; composite cell construct; ds.</p> <p>Rattus sp.</p>				
Key	Location/Qualifiers			
FT	1..36			
FT	/*tag= a			
FT	/partial			
FT	/number= 1			
FT	37..101			
FT	/*tag= b			
FT	/number= 2			
FT	48..317			
FT	/*tag= c			
FT	/product= "Rat amelogenin protein"			
FT	/transl_except= (pos:69..71, aa:Gly)			
FT	48..95			
FT	/*tag= d			
FT	96..314			
FT	/*tag= e			
FT	/label= Mature_rat_amelogenin_protein			
FT	102..149			
FT	/*tag= f			
FT	/number= 3			
FT	150..191			
FT	/*tag= g			
FT	/number= 4			
FT	192..236			
FT	/*tag= h			
FT	/number= 5			
FT	237..311			
FT	/*tag= i			
FT	/number= 6			
FT	/note= "Comprises of gene segments 6a, b, c and d"			
FT	312..317			
FT	/*tag= j			
FT	/number= 7			
FT	/note= "Includes the stop codon"			

Query	Match	28.1%	Score 40.2	DB 3	Length 476
Best Local Similarity	93.3%	Pred. No. 0.0014			
Matches 42	Conservative 0	Mismatches 3	Indels 0	Gaps 0	
DB	59 GATTTGTTGGCCTGCTCCGAGGAGAGCCTTGATGCGCGT 45				
	59 GATCTTGTTGGCCTGCTCCGAGGAGAGCCTTGATGCGCGT 103				
RESULT 7					
ID	ADB59026				
	ADB59026 standard; DNA; 556 BP.				
AC	ADB59026;				
DT	04-DEC-2003 (first entry)				
DE	Toxicity-related gene, SEQ ID 4052.				
OS	Toxic; toxin; gene expression profile; hepatotoxicity; liver;				
PA	drug screening; toxicity assay; ds.				
PN	Unidentified.				
PP	MO2003064624-A2.				
PS	07-AUG-2003.				
PT	31-JAN-2003; 2003WO-US003194.				
PF	31-JAN-2002; 2002US-00060087.				
PI	15-MAR-2002; 2002US-0364045P.				
PR	15-MAR-2002; 2002US-0364055P.				
PP	30-DEC-2002; 2002US-0438643P.				
PP	(GENE-) GENE LOGIC INC.				

XX Mendrick D, Porter M, Johnson K, Higgs B, Caecle A, Elashoff M;  
PI WPI, 2003-689530/65.  
XX Predicting a toxic effect of a compound, useful in identifying toxicity  
PT markers in liver tissues or cells for drug screening and toxicity assays,  
PT comprises preparing gene expression profile of tissue or cells exposed to  
PT the compound.  
XX Claim 1; SEQ ID NO 4052; 1156bp; English.  
XX The present invention relates to a method for predicting a toxic effect  
CC of a compound. The method comprises preparing a gene expression profile  
CC of a tissue or cell sample exposed to the compound, and comparing the  
CC gene expression profile to a database comprising SEQ ID 1-4925, where  
CC differential expression of the gene indicates at least one toxic effect.  
CC The method is useful for predicting at least one toxic effect of a  
CC compound, predicting hepatotoxicity or the progression of a toxic effect  
CC of a compound, identifying an agent that modulates the onset or  
CC progression of a toxic response, predicting the cellular pathways that a  
CC compound modulates in a cell, and identifying an agent that modulates at  
CC least one activity of a protein. The method and compositions of the  
CC present invention using a database of genes having liver toxin-induced  
CC differential expression, are useful in identifying toxicity markers in  
CC liver tissues or cells for drug screening and toxicity assays. Note: The  
CC sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
SQ Sequence 556 BP; 121 A; 183 C; 114 G; 119 T; 0 U; 19 Other;  
Query Match 27.4%; Score 39.2; DB 10; Length 556;  
Best Local Similarity 91.1%; Pred. No. 0.0032;  
Matches 41; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
OY 1 GATTGTGCTGCGCTCCTGGAGAGAGCCTTGTATGCCCGT 45  
Db 33 GATCTGTGCTGCGCTCCTGGAGAGAGCCTTGTATGCCCGT 77  
RESULT 8  
ADBS3782 standard; DNA; 556 BP.  
XX ADBS3782;  
AC XX  
DT 04-DEC-2003 (first entry)  
XX  
DE Primary rat hepatocyte toxicity modelling related gene SEQ ID NO:4324.  
XX  
KM toxic effect; gene expression profile; hepatotoxicity; diagnostic marker;  
KM toxicity marker; toxicity progression; drug screening;  
KM primary rat hepatocyte toxicity modelling; gene; ds.  
XX  
OS Rattus norvegicus.  
XX  
PN WO2003065993-A2.  
PD 14-AUG-2003.  
XX  
PF 04-FEB-2003; 2003MO-US003482.  
XX  
PR 04-FEB-2002; 2002US-0353171P.  
PR 13-MAR-2002; 2002US-0363534P.  
PR 08-APR-2002; 2002US-0370248P.  
PR 10-APR-2002; 2002US-0371134P.  
PR 10-APR-2002; 2002US-0371135P.  
PR 10-APR-2002; 2002US-0371150P.  
PR 11-APR-2002; 2002US-0371413P.  
PR 19-APR-2002; 2002US-0373601P.  
PR 19-APR-2002; 2002US-0373602P.  
PR 22-APR-2002; 2002US-0374139P.  
XX

PR 08-MAY-2002; 2002US-0378370P.  
PR 09-MAY-2002; 2002US-0378652P.  
PR 09-MAY-2002; 2002US-0378653P.  
PR 09-MAY-2002; 2002US-0378655P.  
PR 09-JUL-2002; 2002US-0394230P.  
PR 09-JUL-2002; 2002US-0394253P.  
PR 04-SEP-2002; 2002US-0407688P.  
PR 28-JAN-2003; 2003US-0442900P.  
XX  
PA (GENE-) GENE LOGIC INC.  
XX  
PI Mendrick D, Porter M, Johnson K, Higgs B, Caecle A, Orr M;  
PI Elashoff M;  
XX WPI; 2003-731472/69.  
XX  
XX Determining if a compound induces a toxic effect on a tissue or cell, for  
PT identifying hepatotoxic compounds, comprises comparing a gene expression  
PT profile of a tissue or cell sample to a database of Tox mean and non-Tox  
PT mean values.  
XX  
XX Claim 44; SEQ ID NO 4324; 874bp; English.  
PS  
XX The present invention describes a method for determining whether a  
CC compound induces a toxic effect on a tissue or cell. The method comprises  
CC preparing a gene expression profile of a tissue or cell sample exposed to  
CC the compound, and comparing the gene expression profile to a database  
CC comprising data or information on the Tox mean and non-Tox mean value.  
CC The method is useful for predicting or identifying at least one toxic  
CC effect, particularly hepatotoxicity, of a test or unknown compound. The  
CC genes listed in the specification are useful as diagnostic or toxicity  
CC markers for the prediction or identification of the physiological state  
CC of tissue or cell sample that has been exposed to a compound, or to  
CC identify or predict the toxic effects of a compound or an agent. These  
CC may also be used as markers for monitoring toxicity progression or for  
CC drug screening. The present sequence represents a primary rat hepatocyte  
CC toxicity modelling related gene sequence from the present invention.  
XX  
SQ Sequence 556 BP; 121 A; 183 C; 114 G; 119 T; 0 U; 19 Other;  
Query Match 27.4%; Score 39.2; DB 10; Length 556;  
Best Local Similarity 91.1%; Pred. No. 0.0032;  
Matches 41; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
OY 1 GATTGTGCTGCGCTCCTGGAGAGAGCCTTGTATGCCCGT 45  
Db 33 GATCTGTGCTGCGCTCCTGGAGAGAGCCTTGTATGCCCGT 77  
RESULT 9  
AAL4111 standard; DNA; 722 BP.  
XX AAL4111;  
AC XX  
XX 16-OCT-2002 (first entry)  
DT  
DE gAML related Y-chromosome DNA sequence.  
XX  
KM Goat embryo sexual identification technique; goat amelogenin gene; gAML;  
KM sex-specific; gene; ds; Y-chromosome.  
XX  
OS Capra hircus.  
XX  
XX Key Location/Qualifiers  
FH CDS 35..658  
FT /\*tag= a  
FT /product= "Y-chromosome protein"  
PN TW454013-A.  
PD 11-SEP-2001.  
XX

```
PF 10-NOV-1999; 99TW-00119616.
XX
XX 10-NOV-1999; 99TW-00119616.
XX (CHEN/) CHEN C.
XX (JANG/) JANG J.
XX (WENG/) WENG T.
XX (JENG/) JENG D.
XX
XX Chen C, Jang J, Weng T, Jeng D;
XX WPI; 2002-442016/47.
XX P-PSDB; AAO22534.
XX
XX Sex-specific sequence of goat amelogenin gene, useful for embryo sexual
XX identification, comprises high sensitivity even using single white blood
XX cell or cleavage c.
XX
XX Disclosure; Page 28; 35pp; Chinese.
XX
XX The invention relates to a goat embryo sexual identification technique
XX with high efficiency, sensitivity and repeatability. This technique
XX involves separately cloning and sequencing the coding regions and the
XX introns of the goat amelogenin gene (gAML) on the goat chromosomes. The
XX results indicate that there are sex-specific sequences in the fifth
XX intron of the gene. The major characteristics according to the present
XX invention include high sensitivity, applicable in sex identification even
XX only using a single white blood cell or a single cleavage cell of
XX blastula; high diagnostic efficiency, capable of identifying hundreds of
XX goat embryo in 3 hours; simple operation procedures without complicated
XX steps of DNA extraction and need no additional control group intron; and
XX can be applied on different species of goats. This polynucleotide
XX sequence represents a gAML related X-chromosome DNA sequence of the
XX invention
XX
XX Sequence 722 BP; 173 A; 264 C; 140 G; 145 T; 0 U; 0 Other;
XX
XX Query Match 27.1%; Score 38.8; DB 6; Length 722;
XX Best Local Similarity 95.2%; Pred. No. 0.0048;
XX Matches 40; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX 1 GATTTTGTTGCTGCTCTCTGGAGCAGCCTTGTATGCC 42
XX 46 GATTTTGTTGCTGCTCTCTGGAGCAGCCTTGTATGCC 87
XX
XX RESULT 10
XX AAL41110
XX ID AAL41110 standard; DNA; 752 BP.
XX
XX AAL41110;
XX
XX 16-OCT-2002 (first entry)
XX
XX gAML related X-chromosome DNA sequence.
XX
XX Goat embryo sexual identification technique; goat amelogenin gene; gAML;
XX sex-specific; gene; db; X-chromosome.
XX
XX Capra hircus.
XX
XX Key Location/Qualifiers
XX FH 35..658
XX FT /*tag= a
XX FT /product= "X-chromosome protein"
XX
XX TW454013-A.
XX
XX 11-SEP-2001.
XX
XX 10-NOV-1999; 99TW-00119616.
XX
XX 10-NOV-1999; 99TW-00119616.
XX
XX 10-NOV-1999; 99TW-00119616.
```

```
XX
XX (CHEN/) CHEN C.
XX (JANG/) JANG J.
XX (WENG/) WENG T.
XX (JENG/) JENG D.
XX
XX Chen C, Jang J, Weng T, Jeng D;
XX WPI; 2002-442016/47.
XX P-PSDB; AAO22534.
XX
XX Sex-specific sequence of goat amelogenin gene, useful for embryo sexual
XX identification, comprises high sensitivity even using single white blood
XX cell or cleavage c.
XX
XX Disclosure; Page 28; 35pp; Chinese.
XX
XX The invention relates to a goat embryo sexual identification technique
XX with high efficiency, sensitivity and repeatability. This technique
XX involves separately cloning and sequencing the coding regions and the
XX introns of the goat amelogenin gene (gAML) on the goat chromosomes. The
XX results indicate that there are sex-specific sequences in the fifth
XX intron of the gene. The major characteristics according to the present
XX invention include high sensitivity, applicable in sex identification even
XX only using a single white blood cell or a single cleavage cell of
XX blastula; high diagnostic efficiency, capable of identifying hundreds of
XX goat embryo in 3 hours; simple operation procedures without complicated
XX steps of DNA extraction and need no additional control group intron; and
XX can be applied on different species of goats. This polynucleotide
XX sequence represents a gAML related X-chromosome DNA sequence of the
XX invention
XX
XX Sequence 752 BP; 184 A; 271 C; 152 G; 145 T; 0 U; 0 Other;
XX
XX Query Match 27.1%; Score 38.8; DB 6; Length 752;
XX Best Local Similarity 95.2%; Pred. No. 0.0049;
XX Matches 40; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX 1 GATTTTGTTGCTGCTCTCTGGAGCAGCCTTGTATGCC 42
XX 46 GATTTTGTTGCTGCTCTCTGGAGCAGCCTTGTATGCC 87
XX
XX RESULT 11
XX ADM80836
XX ID ADM80836 standard; cDNA; 318 BP.
XX
XX ADM80836;
XX
XX 03-JUN-2004 (first entry)
XX
XX Human CADEC23 encoding cDNA SEQ ID NO:65.
XX
XX human; cell adhesion and extracellular matrix protein; CADEC23;
XX neuroprotective; cytoskeletal; anorectic; immune disorder;
XX neurological disorder; developmental disorder;
XX connective tissue disorder; cell proliferative disorder; cancer; obesity;
XX Tangle disease; gene; ss.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX FH 36..260
XX FT /*tag= a
XX FT /product= "CADEC23"
XX
XX WO2004015396-A2.
XX
XX 19-FEB-2004.
XX
XX 12-AUG-2003; 2003WO-US025418.
XX
XX 13-AUG-2002; 2002US-0403781P.
```

PR 30-AUG-2002; 2002US-0407034P.  
PR 13-SEP-2002; 2002US-0410566P.  
PR 24-SEP-2002; 2002US-0413482P.  
PR 25-SEP-2002; 2002US-0413890P.  
PR 08-NOV-2002; 2002US-0424904P.  
PR 13-NOV-2002; 2002US-0426222P.  
XX  
PA (INCYTE) INCYTE CORP.  
XX  
PI Ellicott VS, Khare R, Emerling BM, Kable AE, Tran UK, Jin P,  
PI Becha SD, Margolis JP, Swarnakar A, Chawla NK, Ramkumar J,  
PI Hafelia AJA, Lee SY, Jiang X, Jackson AA, Richardson TW, Blake JJ,  
PI Wang JT, Chien D, Yang YG;  
XX WPI; 2004-191795/18.  
DR P-PSDB; ADM60794.  
XX  
PT New cell adhesion and extracellular matrix proteins, useful in  
PT diagnosing, treating and preventing immune, neurological, developmental,  
PT connective tissue and cell proliferative disorders including cancer.  
XX  
PS Claim 5; SEQ ID NO 65; 272bp; English.  
XX  
CC The present sequence encodes a human cell adhesion and extracellular  
CC matrix protein designated CADECM. CADECM sequences has neuroprotective,  
CC cytoskeletal and anorectic activities. The CADECM polypeptides and  
CC polynucleotides are useful in diagnosing, treating and preventing immune,  
CC neurological, developmental, connective tissue and cell proliferative  
CC disorders including cancer, e.g. breast, prostate, ovarian, lung or colon  
CC cancer, obesity and Tangle disease.  
XX  
SQ Sequence 318 BP; 87 A; 81 C; 75 G; 75 T; 0 U; 0 Other;  
Query Match 26.0%; Score 37.2; DB 12; Length 318;  
Best Local Similarity 92.9%; Pred. No. 0.013;  
Matches 39; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 GATTGTGTTGCTGCTCTGCGAGACAGCTTTGCTATGCC 42  
DB 47 GATTTATTGCTGCTCTGCGAGACAGCTTTGCTATGCC 88  
RESULT 12  
AAZ07020  
ID AAZ07020 standard; DNA; 750 BP.  
XX  
AC AAZ07020;  
XX  
DT 15-NOV-1999 (first entry)  
XX  
DE Amelogenin X nucleotide sequence.  
XX  
KW Amelogenin X; Amelogenin Y; genotype; 5' nuclease amplification;  
KW multiple allelic site; apolipoprotein E; apoB; coronary artery disease;  
KW Alzheimer's disease; ds.  
XX  
OS Unidentified.  
XX  
PN WO940226-A2.  
XX  
PD 12-AUG-1999.  
XX  
PF 08-JAN-1999; 99WO-US000499.  
XX  
PR 04-FEB-1998; 98US-00018595.  
XX  
PA (PERKIN) PERKIN-ELMER CORP.  
XX  
PI Liyak KJ, Goodsaiaid F;  
XX WPI; 1999-539985/45.  
DR  
PT 5' nuclease amplification assay using fluorescence-quencher probes for

PT determination of a genotype at multiple allelic sites.  
XX  
PS Disclosure; Fig 10; 95pp; English.  
XX  
CC The present invention describes first and second sets of fluorescer-  
CC quencher probes used simultaneously in a 5' nuclease assay to identify  
CC which members of a first or second set of substantially homologous  
CC sequences are present in a DNA sample. The method can be used to genotype  
CC a sample of genomic DNA at two or more different allelic sites.  
CC Generating a fluorescence spectrum and signature for each genotype, which  
CC uniquely reflects the assay's inherent inefficiency for that genotype  
CC given the particular conditions, probes and primers used, the genotype of  
CC unknown sequences can be determined. The assay was shown to be useful for  
CC determining apob genotypes. The assay can be used as a diagnostic tool  
CC for assessing the risk for coronary artery disease and/or late-onset  
CC Alzheimer's disease. Using the 5' nuclease assay of the invention it is  
CC possible to determine a genotype at two or more allelic sites in a single  
CC reaction. This approach is much faster than previous approaches to  
CC genotyping genes having two or more allelic sites, such as the  
CC apolipoprotein E gene. A key advantage of the method for determining the  
CC genotype of a sample of DNA at multiple allelic sites is that it does not  
CC rely on 5' nuclease assay working with 100% efficiency to distinguish  
CC between substantially homologous sequences such as alleles. The present  
CC sequence represent the nucleotide sequence for amelogenin X, which is  
XX used in the exemplification of the present invention  
SQ Sequence 750 BP; 191 A; 260 C; 146 G; 153 T; 0 U; 0 Other;  
Query Match 26.0%; Score 37.2; DB 2; Length 750;  
Best Local Similarity 92.9%; Pred. No. 0.018;  
Matches 39; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 GATTGTGTTGCTGCTCTGCGAGACAGCTTTGCTATGCC 42  
DB 80 GATTTATTGCTGCTCTGCGAGACAGCTTTGCTATGCC 121  
RESULT 13  
AAZ07018  
ID AAZ07018 standard; DNA; 793 BP.  
XX  
AC AAZ07018;  
XX  
DT 15-NOV-1999 (first entry)  
XX  
DE Amelogenin X nucleotide sequence.  
XX  
KW Amelogenin X; Amelogenin Y; genotype; 5' nuclease amplification;  
KW multiple allelic site; apolipoprotein E; apoB; coronary artery disease;  
KW Alzheimer's disease; ss.  
XX  
OS Unidentified.  
XX  
PN WO940226-A2.  
XX  
PD 12-AUG-1999.  
XX  
PF 08-JAN-1999; 99WO-US000499.  
XX  
PR 04-FEB-1998; 98US-00018595.  
XX  
PA (PERKIN) PERKIN-ELMER CORP.  
XX  
PI Liyak KJ, Goodsaiaid F;  
XX WPI; 1999-539985/45.  
DR  
PT 5' nuclease amplification assay using fluorescence-quencher probes for  
PT determination of a genotype at multiple allelic sites.  
XX  
PS Disclosure; Fig 8A; 95pp; English.  
XX  
PT The present invention describes first and second sets of fluorescer-

CC	quencher probes used simultaneously in a 5' nuclease assay to identify
CC	sequences are present in a DNA sample. The method can be used to genotype
CC	a sample of genomic DNA at two or more different allelic sites.
CC	Generating a fluorescence spectrum and signature for each genotype, which
CC	uniquely reflects the assay's inherent inefficiency for that genotype
CC	given the particular conditions, probes and primers used, the genotype of
CC	unknown sequences can be determined. The assay was shown to be useful for
CC	determining apoc genotypes. The assay can be used as a diagnostic tool
CC	for assessing the risk for coronary artery disease and/or late-onset
CC	Alzheimer's disease. Using the 5' nuclease assay of the invention it is
CC	possible to determine a genotype at two or more allelic sites in a single
CC	reaction. This approach is much faster than previous approaches to
CC	genotyping genes having two or more allelic sites, such as the
CC	apolipoprotein E gene. A key advantage of the method for determining the
CC	genotype of a sample of DNA at multiple allelic sites is that it does not
CC	rely on 5' nuclease assay working with 100% efficiency to distinguish
CC	between substantially homologous sequences such as alleles. The present
CC	sequence represent the nucleotide sequence for ameloogenin X, which is
CC	used in the exemplification of the present invention
SQ	Sequence 793 BP; 215 A; 266 C; 150 G; 162 T; 0 U; 0 Other;
QY	Query Match 26.0%; Score 37.2; DB 2; Length 793;
D8	Best Local Similarity 92.9%; Pred.No.0.018; Indels 0; Gaps 0;
	Matches 39; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Db	1 GATTTTGTTCCTGCGCCTCTGGAGACAGCCTTGCTATGCC 42       80 GATTTAATTTCCTGCGCCTCTGGAGACAGCTTTGCATGCG 121
RESULT 14	
AAZ07019	standard; DNA; 802 BP.
XX	AAZ07019;
AC	
XX	15-NOV-1999 (first entry)
DT	
XX	Amelogenin Y nucleotide sequence.
DE	
XX	Amelogenin X; Amelogenin Y; genotype; 5' nuclease amplification;
KM	multiple allelic site; apolipoprotein E; apoE; coronary artery disease;
KW	Alzheimer's disease; se.
KX	
OS	Unidentified.
XX	
PN	WO9940226-A2.
PD	
XX	12-AUG-1999.
PJ	
XX	08-JAN-1999; 99WO-US0000499.
PF	
XX	04-FEB-1998; 98US-00018595.
PR	
XX	(PEKE ) PERKIN-ELMER CORP.
PA	
PI	Livak KJ, Goodesaid P;
XX	
DR	WPI; 1999-539985/45.
PT	
XX	5' nuclease amplification assay using fluorescence-quencher probes for
XX	determination of a genotype at multiple allelic sites.
XX	
PS	Disclosure; Fig 8B; 95BP; English.
CC	The present invention describes first and second sets of fluorescer-
CC	quencher probes used simultaneously in a 5' nuclease assay to identify
CC	which members of a first or second set of substantially homologous
CC	sequences are present in a DNA sample. The method can be used to genotype
CC	a sample of genomic DNA at two or more different allelic sites.
CC	Generating a fluorescence spectrum and signature for each genotype, which

	CC	uniquely reflects the assay's inherent inefficiency for that genotype
	CC	or given the particular conditions, probes and primers used, the genotype
	CC	of unknown sequences can be determined. The assay was shown to be useful for
	CC	determining Apoe genotypes. The assay can be used as a diagnostic tool
	CC	for assessing the risk for coronary artery disease and/or late-onset
	CC	Alzheimer's disease. Using the 5' nuclease assay of the invention it is
	CC	possible to determine a genotype at two or more allelic sites in a single
	CC	reaction. This approach is much faster than previous approaches to
	CC	genotyping genes having two or more allelic sites, such as the
	CC	Apolipoprotein E gene. A key advantage of the method for determining the
	CC	genotype of a sample of DNA at multiple allelic sites is that it does not
	CC	rely on 5' nuclease assay working with 100% efficiency to distinguish
	CC	between substantially homologous sequences such as alleles. The present
	CC	sequence represent the nucleotide sequence for amelogenin Y, which is
	CC	used in the exemplification of the present invention
SQ		Sequence 802 BP; 216 A; 258 C; 155 G; 173 T; 0 U; 0 Other;
Query Match		24.9%; Score 35.6; DB 2; Length 802;
Best Local Similarity		90.5%; Pred.No. 0.065;
Matches	38; Conservative	0; Mismatches 4; Indels 0; Gaps 0
Dt		1 GATTGTGGCTGCGTCTTGAGACGCCTTTGCATGCGC 42       80 GATTTGTTTCCTCTTGCGAGACGCTTTGCATGCGC 121
RESULT 15		
ID	ADQ22977	ID ADQ22977 standard; DNA; 852 BP.
AC	ADO22977;	
DT	26-AUG-2004	(first entry)
DE		Human soft tissue sarcoma-upregulated DNA - SEQ ID 5797.
KW	soft tissue sarcoma; cyrostatic; gene therapy; vaccine; screening; human; dr.	
XX	Homo sapiens.	
XX	WO2004048938-A2.	
XX	10-JUN-2004.	
Pf	26-NOV-2003; 2003MO-USO38193.	
PR	26-NOV-2002; 2002US-0429739P.	
PA	(PROT-) PROTEIN DESIGN LABS INC.	
Ft	Aziz N, Ginsburg WM, Zlotnick A;	
XX	WPJ; 2004-441208/41.	
PT		Early detection of soft tissue sarcoma comprises determining expression of a gene in a first soft tissue sample and a normal soft tissue sample and comparing the gene expression, also useful in treating soft tissue sarcoma.
Example 2; SEQ ID NO 5797; 210pp; English.		
The invention relates to a novel method for detecting soft tissue sarcomama		
which comprises obtaining a first soft tissue sample from an individual		
and a normal soft tissue sample from the same or different individual,		
determining the expression of a gene in both samples and comparing the		
expression of the gene in both soft tissue samples, where a higher level		
of protein expression in the first soft tissue sample indicates the		
presence of soft tissue sarcoma. The method of the invention has		
cytostatic applications and may be useful for detecting soft tissue		
sarcoma, possibly via gene therapy or vaccine production. The nucleic		
acid sequences may be useful in diagnostic and screening applications.		





**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 8, 2005, 13:03:55 ; Search time 38.6217 Seconds  
(without alignments)  
2631.757 Million cell updates/sec

Title: US-10-754-437-10

Perfect score: 143

Sequence: 1 gattctgttgcctgcctcc.....gtttaagacagtaacagatc 143

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 824507 seqs, 355394441 residues

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:\*

- 1: /cgn2\_6/prodata/1/ina/5A\_COMB.seq:\*
- 2: /cgn2\_6/prodata/1/ina/5B\_COMB.seq:\*
- 3: /cgn2\_6/prodata/1/ina/6A\_COMB.seq:\*
- 4: /cgn2\_6/prodata/1/ina/6B\_COMB.seq:\*
- 5: /cgn2\_6/prodata/1/ina/PCTUS\_COMB.seq:\*
- 6: /cgn2\_6/prodata/1/ina/backfillseq1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	40.2	28.1	476	4 US-09-744-128-3	Sequence 3, Appl
2	37.2	26.0	500	2 US-09-018-595B-3	Sequence 3, Appl
3	37.2	26.0	500	3 US-09-324-709A-3	Sequence 3, Appl
4	37.2	26.0	793	2 US-09-018-595B-1	Sequence 1, Appl
5	37.2	26.0	793	3 US-09-324-709A-1	Sequence 1, Appl
6	35.6	24.9	802	2 US-09-018-595B-2	Sequence 2, Appl
7	35.6	24.9	802	3 US-09-324-709A-2	Sequence 2, Appl
8	30	21.0	1284	4 US-09-328-352-2004	Sequence 2004, Ap
9	30	21.0	1230025	4 US-09-198-452A-1	Sequence 1, Appl
10	29.6	20.7	116592	4 US-09-818-512-3	Sequence 3, Appl
11	29.4	20.6	1431	3 US-09-054-782-1	Sequence 1, Appl
12	29.4	20.6	1431	4 US-09-627-676-1	Sequence 1, Appl
13	29.4	20.6	1431	4 US-09-054-272-15	Sequence 15, Appl
14	29.4	20.6	1431	6 5466783-1	Patent No. 5466783
15	29.4	20.6	45175	3 US-09-453-702B-116	Sequence 116, App
16	29	20.3	3568	3 US-09-103-510B-1	Sequence 1, Appl
17	29	20.3	3568	4 US-09-492-639A-1	Sequence 1, Appl
18	29	20.3	3568	4 US-09-491-971A-1	Sequence 1, Appl
19	29	20.3	3568	4 US-09-491-970A-1	Sequence 1, Appl
20	28.6	20.0	895	4 US-09-270-767-31527	Sequence 31527, A
21	28.6	20.0	895	4 US-09-270-767-26797	Sequence 26797, A
22	28.6	20.0	1514	4 US-09-270-767-15199	Sequence 15199, A
23	28.6	20.0	1629	4 US-09-270-767-11252	Sequence 11252, A
24	28.2	19.7	1410	4 US-09-861-451A-13	Sequence 13, Appl
25	27.8	19.4	792	4 US-09-710-279-203	Sequence 203, App
26	27.8	19.4	3841	4 US-09-710-279-3380	Sequence 3380, Ap
27	27.6	19.3	3666	4 US-09-710-279-4230	Sequence 4230, Ap

28	27.6	19.3	4066	4 US-09-710-279-4327	Sequence 4327, Ap
29	27.4	19.2	510	4 US-09-134-000C-106	Sequence 106, App
30	27	18.9	1138	4 US-09-270-767-26204	Sequence 26204, A
31	27	18.9	1669	4 US-09-591-466C-1	Sequence 1, Appl
32	27	18.9	2123	4 US-09-270-767-10744	Sequence 10744, A
33	26.8	18.7	1914	4 US-09-248-796A-4911	Sequence 4911, Ap
34	26.8	18.7	5966	4 US-08-956-171E-22	Sequence 22, Appl
35	26.8	18.7	5966	4 US-08-781-986A-22	Sequence 22, Appl
36	26.6	18.6	9262	4 US-09-566-921-74	Sequence 74, Appl
37	26.4	18.5	528	4 US-09-248-796A-6060	Sequence 6060, Ap
38	26.4	18.5	2070	4 US-09-540-236-1766	Sequence 1766, Ap
39	26.4	18.5	89047	4 US-09-596-002-34	Sequence 34, Appl
40	26.2	18.3	804	3 US-09-134-001C-2205	Sequence 2205, Ap
41	26	18.2	936	4 US-09-489-039A-6691	Sequence 6691, Ap
42	26	18.2	1032	4 US-09-248-796A-8331	Sequence 8331, Ap
43	26	18.2	1511	3 US-08-847-065-18	Sequence 18, Appl
44	26	18.2	1511	4 US-09-829-382-18	Sequence 18, Appl
45	25.8	18.0	444	4 US-09-543-681A-1204	Sequence 1204, Ap

#### ALIGNMENTS

RESULT 1  
US-09-744-128-3  
Sequence 3, Application US/09744128  
Patent No. 6677306  
GENERAL INFORMATION:  
APPLICANT: Vels et al.  
TITLE OF INVENTION: Chondrogenic and Osteogenic Inducing Molecule  
FILE REFERENCE: 2766/36983  
CURRENT APPLICATION NUMBER: US/09/744,128  
PRIOR FILING DATE: 2001-05-16  
PRIOR APPLICATION NUMBER: PCT/US99/17342  
PRIOR FILING DATE: 1999-07-29  
PRIOR APPLICATION NUMBER: 60/094,489  
NUMBER OF SEQ ID NOS: 17  
SOFTWARE: Patentln 3.1  
SEQ ID NO 3  
LENGTH: 476  
TYPE: DNA  
ORGANISM: Artificial sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: PCR Product  
US-09-744-128-3  
Query Match 28.1%; Score 40.2; DB 4; Length 476;  
Best Local Similarity 93.3%; Pred. No. 6.8e-05;  
Matches 42; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 GATTTGTTGCTGCTGCTGAGACAGCTTGTATGCCGT 45  
DB 59 GATCTGTTGCTGCTGCTGAGACAGCTTGTATGCCCT 103  
RESULT 2  
US-09-018-595B-3  
Sequence 3, Application US/09018595B  
Patent No. 5962233  
GENERAL INFORMATION:  
APPLICANT: Perkin-Elmer Corporation,  
TITLE OF INVENTION: Applied Biosystems Division  
DETERMINATION OF GENOTYPE OF  
TITLE OF INVENTION: AMPLIFICATION PRODUCTS AT MULTIPLE ALLELIC SITES  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSER: David J. Weitz,  
ADDRESSER: Wilson Somaini Goodrich & Rosati  
STREET: 650 Page Mill Road  
CITY: Palo Alto  
STATE: California  
COUNTRY: USA

ZIP: 94304-1050  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch diskette  
COMPUTER: IBM compatible  
OPERATING SYSTEM: Microsoft Windows 95  
SOFTWARE: Wordperfect for windows 6.0,  
SOFTWARE: ASCII (DOS) TEXT format  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/018,595B  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: David J. Weitz  
REGISTRATION NUMBER: 38,362  
REFERENCE/DOCKET NUMBER: PEIM-744  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 493-9300  
TELEFAX: (650) 493-6811  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 500 nucleotides  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-09-018-595B-3

Query Match 26.0%; Score 37.2; DB 2; Length 500;  
Best Local Similarity 92.9%; Pred. No. 0.00086;  
Matches 39; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 GATTTGTTGCTGCTCCTCGGAGACGCTTGTGATGCC 42  
Db 80 GATTTATTGCTGCTCCTCGGAGACGCTTGTGATGCC 121

RESULT 3  
US-09-324-709A-3  
Sequence 3, Application US/09324709A  
Patent No. 6154707  
GENERAL INFORMATION:  
APPLICANT: Perkin-Elmer Corporation,  
APPLICANT: Applied Biosystems Division  
TITLE OF INVENTION: FLUORESCENCE GENOTYPING AT MULTIPLE ALLELIC  
TITLE OF INVENTION: SITES  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David J. Weitz,  
ADDRESSEE: Wilson Sonsini Goodrich & Rosati  
STREET: 650 Page Mill Road  
CITY: Palo Alto  
STATE: California  
COUNTRY: USA  
ZIP: 94304-1050  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch diskette  
COMPUTER: IBM compatible  
OPERATING SYSTEM: Microsoft Windows 95  
SOFTWARE: Wordperfect for windows 6.0,  
SOFTWARE: ASCII (DOS) TEXT format  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/324,709A  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: David J. Weitz  
REGISTRATION NUMBER: 38,362  
REFERENCE/DOCKET NUMBER: 16842-758

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 493-9300  
TELEFAX: (650) 493-6811  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 500 nucleotides  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-09-324-709A-3

Query Match 26.0%; Score 37.2; DB 3; Length 500;  
Best Local Similarity 92.9%; Pred. No. 0.00086;  
Matches 39; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 GATTTGTTGCTGCTCCTCGGAGACGCTTGTGATGCC 42  
Db 80 GATTTATTGCTGCTCCTCGGAGACGCTTGTGATGCC 121

RESULT 4  
US-09-018-595B-1  
Sequence 1, Application US/09018595B  
Patent No. 5962233  
GENERAL INFORMATION:  
APPLICANT: Perkin-Elmer Corporation,  
APPLICANT: Applied Biosystems Division  
TITLE OF INVENTION: DETERMINATION OF GENOTYPE OF  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David J. Weitz,  
ADDRESSEE: Wilson Sonsini Goodrich & Rosati  
STREET: 650 Page Mill Road  
CITY: Palo Alto  
STATE: California  
COUNTRY: USA  
ZIP: 94304-1050  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch diskette  
COMPUTER: IBM compatible  
OPERATING SYSTEM: Microsoft Windows 95  
SOFTWARE: Wordperfect for windows 6.0,  
SOFTWARE: ASCII (DOS) TEXT format  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/018,595B  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: David J. Weitz  
REGISTRATION NUMBER: 38,362  
REFERENCE/DOCKET NUMBER: PEIM-744  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 493-9300  
TELEFAX: (650) 493-6811  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 793 nucleotides  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-018-595B-1

Query Match 26.0%; Score 37.2; DB 2; Length 793;  
Best Local Similarity 92.9%; Pred. No. 0.0011;  
Matches 39; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 GATTTGTTGCTGCTCCTCGGAGACGCTTGTGATGCC 42  
Db 80 GATTTATTGCTGCTCCTCGGAGACGCTTGTGATGCC 121

RESULT 5  
US-09-324-709A-1  
Sequence 1, Application US/09324709A  
Patent No. 6154707  
GENERAL INFORMATION:  
APPLICANT: Perkin-Elmer Corporation,  
APPLICANT: Applied Biosystems Division  
TITLE OF INVENTION: FLUORESCENCE GENOTYPING AT MULTIPLE ALLELIC  
SITES  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: David J. Weltz,  
ADDRESSEE: Wilson Sonaini Goodrich & Rosati  
STREET: 650 Page Mill Road  
CITY: Palo Alto  
STATE: California  
COUNTRY: USA  
ZIP: 94304-1050  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch diskette  
COMPUTER: IBM compatible  
OPERATING SYSTEM: Microsoft Windows 95  
SOFTWARE: Wordperfect for windows 6.0,  
SOFTWARE: ASCII (DOS) TEXT format  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/324,709A  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: David J. Weltz  
REGISTRATION NUMBER: 38,362  
REFERENCE/DOCKET NUMBER: 16842-758  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 493-9300  
TELEFAX: (650) 493-6811  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 793 nucleotides  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-324-709A-1  
Query Match 26.0%; Score 37.2; DB 3; Length 793;  
Best Local Similarity 92.9%; Pred. No. 0.0011; 3; Indels 0; Gaps 0;  
Matches 39; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
Qy 1 GATTTGTTTGCTGCTGCTGAGAGAGCCTTTGCTATGCC 42  
Db 80 GATTATTGCTGCTGCTGAGAGAGCCTTTGCTATGCC 121  
RESULT 6  
US-09-018-595B-2  
Sequence 2, Application US/09018595B  
Patent No. 596223  
GENERAL INFORMATION:  
APPLICANT: Perkin-Elmer Corporation,  
APPLICANT: Applied Biosystems Division  
TITLE OF INVENTION: DETERMINATION OF GENOTYPE OF  
TITLE OF INVENTION: AMPLIFICATION PRODUCTS AT MULTIPLE ALLELIC SITES  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: David J. Weltz,  
ADDRESSEE: Wilson Sonaini Goodrich & Rosati  
STREET: 650 Page Mill Road  
CITY: Palo Alto  
STATE: California

COUNTRY: USA  
ZIP: 94304-1050  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch diskette  
COMPUTER: IBM compatible  
OPERATING SYSTEM: Microsoft Windows 95  
SOFTWARE: Wordperfect for windows 6.0,  
SOFTWARE: ASCII (DOS) TEXT format  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/018,595B  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: David J. Weltz  
REGISTRATION NUMBER: 38,362  
REFERENCE/DOCKET NUMBER: PELM-744  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 493-9300  
TELEFAX: (650) 493-6811  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 802 nucleotides  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-018-595B-2  
Query Match 24.9%; Score 35.6; DB 2; Length 802;  
Best Local Similarity 90.5%; Pred. No. 0.0041; 4; Indels 0; Gaps 0;  
Matches 38; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
Qy 1 GATTTGTTTGCTGCTGCTGAGAGAGCCTTTGCTATGCC 42  
Db 80 GATTATTGCTGCTGCTGAGAGAGCCTTTGCTATGCC 121  
RESULT 7  
US-09-324-709A-2  
Sequence 2, Application US/09324709A  
Patent No. 6154707  
GENERAL INFORMATION:  
APPLICANT: Perkin-Elmer Corporation,  
APPLICANT: Applied Biosystems Division  
TITLE OF INVENTION: FLUORESCENCE GENOTYPING AT MULTIPLE ALLELIC  
SITES  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: David J. Weltz,  
ADDRESSEE: Wilson Sonaini Goodrich & Rosati  
STREET: 650 Page Mill Road  
CITY: Palo Alto  
STATE: California  
COUNTRY: USA  
ZIP: 94304-1050  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch diskette  
COMPUTER: IBM compatible  
OPERATING SYSTEM: Microsoft Windows 95  
SOFTWARE: Wordperfect for windows 6.0,  
SOFTWARE: ASCII (DOS) TEXT format  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/324,709A  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: David J. Weltz  
REGISTRATION NUMBER: 38,362

```
REFERENCE/DOCKET NUMBER: 16842-758
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 493-9300
TELEFAX: (650) 493-6811
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 802 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-324-709A-2

Query Match
Best Local Similarity 90.5%; Pred. No. 0.0041;
Matches 38; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GATTTTGTTCCTGCTGCTGAGAGAGCTTTGCTATGCC 42
DB 80 GATTTTGTTCCTGCTGCTGAGAGAGCTTTGCTATGCC 121

RESULT 8
US-09-328-352-2004/c
Sequence 2004, Application US/09328352
Patent No. 6562958
GENERAL INFORMATION:
APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
FILE REFERENCE: GTC99-03PA
CURRENT APPLICATION NUMBER: US/09/328,352
CURRENT FILING DATE: 1999-06-04
NUMBER OF SEQ ID NOS: 8252
SEQ ID NO 2004
LENGTH: 1284
TYPE: DNA
ORGANISM: Acinetobacter baumannii
US-09-328-352-2004

Query Match
Best Local Similarity 57.4%; Pred. No. 0.56;
Matches 54; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 50 AAAATCCCTGCATATATTCATTTGACAGCTTGAATAAGTCCGCCACAG 109
DB 715 AGAAGAGCTTCACTACCTACCTACCTACCTGACGACACTTTGCGCTTCAA 656
QY 110 TTGTAACTTTAGGCTTTAGACAGTACAGATC 143
DB 655 TTGTAACTTTAGGCTTTGACGACACTTCAAGAC 622

RESULT 9
US-09-198-452A-1/c
Sequence 1, Application US/09198452A
Patent No. 6559294
GENERAL INFORMATION:
APPLICANT: Griffiths, R.
TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention
FILE REFERENCE: 9710-003-999
CURRENT APPLICATION NUMBER: US/09/198,452A
CURRENT FILING DATE: 1998-11-24
NUMBER OF SEQ ID NOS: 6849
SEQ ID NO 1
LENGTH: 1230025
TYPE: DNA
ORGANISM: Chlamydia pneumoniae
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)-(15000)
OTHER INFORMATION: n=a or c or g or t
```

```
NAME/KEY: misc_feature
LOCATION: (15001)..(30000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (30001)..(45000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (45001)..(60000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (60001)..(75000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (75001)..(90000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (90001)..(105000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (105001)..(120000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (120001)..(135000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (135001)..(150000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (150001)..(165000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (165001)..(180000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (180001)..(195000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (195001)..(210000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (210001)..(225000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (225001)..(240000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (240001)..(255000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (255001)..(270000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (270001)..(285000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (285001)..(300000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (300001)..(315000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (315001)..(330000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (330001)..(345000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (345001)..(360000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (360001)..(375000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
```

```
LOCATION: (375001)..(390000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (390001)..(405000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (405001)..(420000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (420001)..(435000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (435001)..(450000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (450001)..(465000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (465001)..(480000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (480001)..(495000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (495001)..(510000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (510001)..(525000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (525001)..(540000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (540001)..(555000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (555001)..(570000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (570001)..(585000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (585001)..(600000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (600001)..(615000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (615001)..(630000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (630001)..(645000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (645001)..(660000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (660001)..(675000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (675001)..(690000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (690001)..(705000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (705001)..(720000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (720001)..(735000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (735001)..(750000)
```

```
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (750001)..(765000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (765001)..(780000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (780001)..(795000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (795001)..(810000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (810001)..(825000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (825001)..(840000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (840001)..(855000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (855001)..(870000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (870001)..(885000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (885001)..(900000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (900001)..(915000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
```

Query Match 21.0%; Score 30; DB 4; Length 1230025;  
Best Local Similarity 55.9%; Pred. No. 15;  
Matches 57; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

Qy 35 GCTATGCCCGTGAATAATATCCCGCATATATTCATTTTCACACAGTTGGAAATAA 94  
Db 980791 GCTATTTCTTTTTCATTTTCGATGATCATATATTTCTTAAGTACTAAGTTTA 980732

Qy 95 AGTTCGCCCCACGAGTTGTACTTAGAGTTTAAGACATA 136  
Db 980731 ATCCTTCGCATCTTCGAAACCATTTGGTTAATAAGCA 980690

RESULT 10  
US-09-818-512-3  
Sequence 3, Application US/09818512  
Patent No. 6537780  
GENERAL INFORMATION:  
APPLICANT: BEASLEY, Ellen et al.  
TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC  
ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES  
FILE REFERENCE: CL001192  
CURRENT APPLICATION NUMBER: US/09/818,512  
CURRENT FILING DATE: 2001-03-28  
NUMBER OF SEQ ID NOS: 4  
SOFTWARE: PaedSeq for Windows Version 4.0  
SEQ ID NO 3  
LENGTH: 116592  
TYPE: DNA  
ORGANISM: Human  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)..(116592)  
OTHER INFORMATION: n = A,T,C or G  
US-09-818-512-3

Query Match 20.7%; Score 29.6; DB 4; Length 116592;  
Best Local Similarity 59.5%; Pred. No. 6.9; Mismatches 34; Indels 0; Gaps 0;  
Matches 50; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

Qy 49 TAAATACCCCTGCATATATTCATTTACACAGCTTGGAAATTAAGTCCCAACCA 108  
Db 15953 TACATATCCCTTGTGATTTGATTTATTAAGTCTTCATTAAGATTTGACTTTA 16012

Qy 109 GTTGTAAGTTAGGTTAGAC 132  
Db 16013 TTGTATAGTTTATTTACGTAC 16036

RESULT 11  
US-09-054-782-1/c  
; Sequence 1, Application US/09054782  
; Patent No. 6171587  
; GENERAL INFORMATION:  
; APPLICANT: Mun, T.C.  
; APPLICANT: Kretzmer, K.K.  
; APPLICANT: Broze, Jr., G.J.  
; TITLE OF INVENTION: DNA clone of human tissue  
; TITLE OF INVENTION: factor inhibitor  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Monsanto/Searle  
; STREET: 700 Chesterfield Parkway No. 6171587th  
; CITY: Chesterfield  
; STATE: MO  
; COUNTRY: USA  
; ZIP: 63198  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/054,782  
; FILING DATE: 03-APR-1998  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/463,602  
; FILING DATE: 05-JUN-1995  
; APPLICATION NUMBER: 08/355,351  
; FILING DATE: 13-DEC-1994  
; APPLICATION NUMBER: 08/093,285  
; FILING DATE: 15-JUL-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bennett, Dennis A  
; REGISTRATION NUMBER: 34,547  
; REFERENCE/DOCKET NUMBER: M 4941-9  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 314-737-6986  
; TELEFAX: 737-6972  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1431 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; ORIGINAL SOURCE:  
; TISSUE TYPE: Human placenta  
; IMMEDIATE SOURCE:  
; LIBRARY: Human placental-derived lambda-P9 clone  
; CLONE: Tissue factor inhibitor (TFI) cDNA  
; FEATURE:  
; NAME/KEY: 5' UTR  
; LOCATION: 1...132  
; OTHER INFORMATION: \*tag=a  
; NAME/KEY: 3' UTR  
; LOCATION: 1045...1431

OTHER INFORMATION: \*tag=b  
; NAME/KEY: CDS  
; LOCATION: 3...104  
; OTHER INFORMATION: \*tag=c  
; PUBLICATION INFORMATION:  
; DOCUMENT NUMBER: EP 318451  
; FILING DATE: 22-JUL-1988  
; PUBLICATION DATE: 31-MAY-1988  
; RELEVANT RESIDUES IN SEQ ID NO:  
US-09-054-782-1

Query Match 20.6%; Score 29.4; DB 3; Length 1431;  
Best Local Similarity 63.4%; Pred. No. 0.97; Mismatches 26; Indels 0; Gaps 0;  
Matches 45; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

Qy 40 GCCCGTAGTAATACCCCTGCATATATTCATTTACACAGCTTGGAAATTAAGTCT 99  
Db 1257 GCCAGTTAATTAATTAAGACCTAGAAATTAAGCAATTTACAGAAAGCAACTCT 1198

Qy 100 GCCCCACCACT 110  
Db 1197 GATACAAAGT 1187

RESULT 12  
US-09-627-676-1/c  
; Sequence 1, Application US/09627676  
; Patent No. 6534276  
; GENERAL INFORMATION:  
; APPLICANT: Broze, George J.  
; APPLICANT: Kretzmer, Kuniko K.  
; APPLICANT: Mun, Tze-Chen  
; TITLE OF INVENTION: Antibodies to human tissue factor  
; TITLE OF INVENTION: inhibitor  
; FILE REFERENCE: SM04941-10-US  
; CURRENT APPLICATION NUMBER: US/09/627,676  
; CURRENT FILING DATE: 2000-07-28  
; PRIOR APPLICATION NUMBER: 09/054,782  
; PRIOR FILING DATE: 1998-04-03  
; PRIOR APPLICATION NUMBER: 08/463,323  
; PRIOR FILING DATE: 1995-06-05  
; PRIOR APPLICATION NUMBER: 08/355,351  
; PRIOR FILING DATE: 1993-07-15  
; PRIOR APPLICATION NUMBER: 07/566,280  
; PRIOR FILING DATE: 1990-08-13  
; PRIOR APPLICATION NUMBER: 07/123,753  
; PRIOR FILING DATE: 1987-11-23  
; PRIOR APPLICATION NUMBER: 07/077,366  
; PRIOR FILING DATE: 1987-07-23  
; SOFTWARE: FastSeq for Windows Version 4.0  
; NUMBER OF SEQ ID NOS: 2  
; SEQ ID NO 1  
; LENGTH: 1431  
; TYPE: DNA  
; ORGANISM: human  
US-09-627-676-1

Query Match 20.6%; Score 29.4; DB 4; Length 1431;  
Best Local Similarity 63.4%; Pred. No. 0.97; Mismatches 26; Indels 0; Gaps 0;  
Matches 45; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

Qy 40 GCCCGTAGTAATACCCCTGCATATATTCATTTACACAGCTTGGAAATTAAGTCT 99  
Db 1257 GCCAGTTAATTAATTAAGACCTAGAAATTAAGCAATTTACAGAAAGCAACTCT 1198

Qy 100 GCCCCACCACT 110  
Db 1197 GATACAAAGT 1187

RESULT 13  
US-09-054-272-15/c  
; Sequence 15, Application US/09054272



Patent No. 6692909  
GENERAL INFORMATION:  
APPLICANT: Lander, Eric S.  
APPLICANT: Daley, George Q.  
APPLICANT: Cargill, Michele  
APPLICANT: Ireland, James S.  
APPLICANT: Rozen, Steven G.  
TITLE OF INVENTION: CODING SEQUENCE POLYMORPHISMS  
TITLE OF INVENTION: IN VASCULAR PATHOLOGY GENES  
NUMBER OF SEQUENCES: 59  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: HAMILTON, BROOK, SMITH & REYNOLDS, P.C.  
STREET: Two Militia Drive  
City: Lexington  
STATE: MA  
COUNTRY: USA  
ZIP: 02173  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows 95  
SOFTWARE: FASTSEQ for Windows Version 2.0b  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/054,272  
FILING DATE: 01-APR-1998  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Granahan, Patricia  
REGISTRATION NUMBER: 32,227  
REFERENCE/DOCKET NUMBER: WH198-05  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 781-861-6240  
TELEFAX: 781-861-9540  
TELEX:  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1431 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: Genomic DNA  
FEATURE:  
NAME/KEY: Coding Sequence  
LOCATION: 133...1044  
OTHER INFORMATION:  
US-09-054-272-15

Query Match 20.6%; Score 29.4; DB 4; Length 1431;  
Best Local Similarity 63.4%; Pred. No. 0.97;  
Matches 45; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 40 GCCCGTAGTAATAACCCCTGCATATATTCAATTTCACAGCTTGAAATAAAGTCT 99  
DB 1257 GCCAGTTAATAATTACAGACTGAGATTAAGCAATTTAACAAGATTGAAAGCAACTCT 1198

QY 100 GCCCCACCACT 110  
DB 1197 GATACAAAGCT 1187

RESULT 14  
5466783-1/c  
Patent No. 5466783  
APPLICANT: Mun, Tze-Chen, Kretzmer, Kuniko K., Broze, George J. Jr.  
TITLE OF INVENTION: HUMAN TISSUE FACTOR INHIBITOR  
NUMBER OF SEQUENCES: 26  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/93,285  
FILING DATE: 15-JUL-1993

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 566,280  
FILING DATE: 13-AUG-1990  
APPLICATION NUMBER: 123,753  
FILING DATE: 23-NOV-1987  
APPLICATION NUMBER: 77,366  
FILING DATE: 23-JUL-1987  
SEQ ID NO: 11:  
LENGTH: 1431  
5466783-1

Query Match 20.6%; Score 29.4; DB 6; Length 1431;  
Best Local Similarity 63.4%; Pred. No. 0.97;  
Matches 45; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 40 GCCCGTAGTAATAACCCCTGCATATATTCAATTTCACAGCTTGAAATAAAGTCT 99  
DB 1257 GCCAGTTAATAATTACAGACTGAGATTAAGCAATTTAACAAGATTGAAAGCAACTCT 1198

QY 100 GCCCCACCACT 110  
DB 1197 GATACAAAGCT 1187

RESULT 15  
US-09-453-702B-116  
Sequence 116, Application US/09453702B  
Patent No. 6365723  
GENERAL INFORMATION:  
APPLICANT: Blattner, Frederick R.  
Burland, Valerie  
Perna, Nicole T.  
Plunkett, Guy  
Welch, Rod  
TITLE OF INVENTION: No. 6365723el Sequences of E. coli O157  
NUMBER OF SEQUENCES: 265  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Charles & Brady  
STREET: 1 South Plinkney Street  
CITY: Madison  
STATE: WI  
COUNTRY: US  
ZIP: 53701-2113  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.44Mb storage  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Word Perfect 8.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/453,702B  
FILING DATE: 03-Dec-1999  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/110,955  
FILING DATE: 04-DEC-1998  
ATTORNEY/AGENT INFORMATION:  
NAME: Seay, Nicholas J.  
REGISTRATION NUMBER: 27386  
REFERENCE/DOCKET NUMBER: 960296, 95017  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (608) 251-5000  
TELEFAX: (608) 251-9166  
INFORMATION FOR SEQ ID NO: 116:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 45175  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
SEQUENCE DESCRIPTION: SEQ ID NO: 116:  
US-09-453-702B-116

Query Match 20.6%; Score 29.4; DB 3; Length 45175;

Best Local Similarity 60.8%; Pred. No. 5.2;  
Matches 48; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 44 GTGAGTAAATACCCCTGCATATATTCATTTCAAGCTTGAAATAAAGTCTGCC 103

Db 82 GTTAGCATCAACCACTTCGAATATTCAGCAGCAACAGGTTAAATAACTGCACCC 141

QY 104 CACCAAGTGGTAATTAG 122

Db 142 GAACAATTGATACGACAG 160

Search completed: February 8, 2005, 22:26:15  
Job time : 41.7328 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using BW model

Run on: February 8, 2005, 16:37:21 ; Search time 192.692 Seconds  
(without alignments)  
4264.119 Million cell updates/sec

Title: US-10-754-437-10

Perfect score: 143  
Sequence: 1 gattctgttcgtcgtcc.....gttaagacagcacaagatc 143

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4300275 seqs, 2872944193 residues

Total number of hits satisfying chosen parameters: 8600550

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Published Applications NA:\*

- 1: /cgn2\_6/prodata/2/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/prodata/2/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/prodata/2/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/prodata/2/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/prodata/2/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/prodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/prodata/2/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/prodata/2/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/prodata/2/pubpna/US09A\_PUBCOMB.seq:\*
- 10: /cgn2\_6/prodata/2/pubpna/US09B\_PUBCOMB.seq:\*
- 11: /cgn2\_6/prodata/2/pubpna/US09C\_PUBCOMB.seq:\*
- 12: /cgn2\_6/prodata/2/pubpna/US09\_NEW\_PUB.seq:\*
- 13: /cgn2\_6/prodata/2/pubpna/US10\_PUBCOMB.seq:\*
- 14: /cgn2\_6/prodata/2/pubpna/US10B\_PUBCOMB.seq:\*
- 15: /cgn2\_6/prodata/2/pubpna/US10C\_PUBCOMB.seq:\*
- 16: /cgn2\_6/prodata/2/pubpna/US10D\_PUBCOMB.seq:\*
- 17: /cgn2\_6/prodata/2/pubpna/US10E\_PUBCOMB.seq:\*
- 18: /cgn2\_6/prodata/2/pubpna/US10\_NEW\_PUB.seq:\*
- 19: /cgn2\_6/prodata/2/pubpna/US11\_NEW\_PUB.seq:\*
- 20: /cgn2\_6/prodata/2/pubpna/US60\_NEW\_PUB.seq:\*
- 21: /cgn2\_6/prodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	75.8	53.0	600	13	US-10-027-632-287050
C 2	75.8	53.0	600	15	US-10-027-632-287050
C 3	74.6	52.2	600	13	US-10-027-632-287049
C 4	74.6	52.2	600	15	US-10-027-632-287049
5	74.6	52.2	259202	18	US-10-723-860-1311
6	40.2	28.1	500	10	US-09-908-975-1137
7	37.2	26.0	500	13	US-10-104-774-3
8	37.2	26.0	500	16	US-10-455-150-3
9	37.2	26.0	793	13	US-10-104-774-1
10	37.2	26.0	793	16	US-10-455-150-1
11	37.2	26.0	793	17	US-10-755-889-605
12	35.6	24.9	802	13	US-10-104-774-2

13	35.6	24.9	802	16	US-10-455-150-2	Sequence 2, Appl1
14	35.6	24.9	852	18	US-10-723-860-5797	Sequence 5797, Ap
C 15	32.6	22.8	680	13	US-10-027-632-218969	Sequence 218969,
C 16	32.6	22.8	680	15	US-10-027-632-218969	Sequence 218969,
17	30.6	21.4	253861	17	US-10-741-601-5611	Sequence 5611, Ap
18	30.6	21.4	261817	13	US-10-087-192-2002	Sequence 2002, Ap
C 19	30.4	21.3	625	13	US-10-027-632-180446	Sequence 180446,
C 20	30.4	21.3	625	13	US-10-027-632-180447	Sequence 180447,
C 21	30.4	21.3	625	13	US-10-027-632-180448	Sequence 180448,
C 22	30.4	21.3	625	15	US-10-027-632-180446	Sequence 180446,
C 23	30.4	21.3	625	15	US-10-027-632-180447	Sequence 180447,
C 24	30.4	21.3	625	15	US-10-027-632-180448	Sequence 180448,
C 25	30.4	21.3	47804	15	US-10-017-161-717	Sequence 717, App
26	30	21.0	47804	15	US-10-292-796-629	Sequence 629, App
27	30	21.0	68140	13	US-10-087-192-1681	Sequence 1681, Ap
C 28	29.8	20.8	1230025	16	US-10-289-762-1	Sequence 1, Appl1
C 29	29.8	20.8	237	9	US-09-864-761-18974	Sequence 18974, A
C 30	29.8	20.8	520	16	US-10-264-049-2155	Sequence 2155, Ap
C 31	29.8	20.8	1513	9	US-09-938-842A-4427	Sequence 4427, Ap
C 32	29.8	20.8	1513	11	US-09-938-842A-4427	Sequence 4427, Ap
C 33	29.8	20.8	60327	16	US-10-052-482-187	Sequence 187, App
C 34	29.6	20.7	251	11	US-09-922-293-2795	Sequence 2795, Ap
C 35	29.6	20.7	116592	9	US-09-818-512-3	Sequence 3, Appl1
36	29.6	20.7	116592	15	US-10-354-065-3	Sequence 3, Appl1
37	29.4	20.6	741	14	US-10-198-846-9976	Sequence 9976, Ap
C 38	29.4	20.6	1431	15	US-10-377-817-1	Sequence 1, Appl1
C 39	29.4	20.6	1431	18	US-10-886-556-1	Sequence 1, Appl1
C 40	29.4	20.6	1612	14	US-10-198-846-9844	Sequence 9844, Ap
C 41	29.4	20.6	2849	9	US-09-880-107-2363	Sequence 2363, Ap
42	29.4	20.6	45175	14	US-10-114-170-116	Sequence 116, App
43	29.4	20.6	167343	9	US-09-962-436-281	Sequence 281, App
44	29.4	20.6	167343	9	US-09-964-824A-273	Sequence 273, App
45	29.4	20.6	2140405	13	US-10-027-632-76212	Sequence 76212, A

#### ALIGNMENTS

RESULT 1  
US-10-027-632-287050/c  
Sequence 287050, US20020198371A1  
Publication No. US20020198371A1  
GENERAL INFORMATION:  
APPLICANT: Wang, David G.  
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
POLYMORPHISMS IN THE HUMAN GENOME  
FILE REFERENCE: 108827.129  
CURRENT APPLICATION NUMBER: US/10/027, 632  
CURRENT FILING DATE: 2002-04-30  
PRIOR APPLICATION NUMBER: US 60/218, 006  
PRIOR FILING DATE: 2000-07-12  
PRIOR APPLICATION NUMBER: US 60/198, 676  
PRIOR FILING DATE: 2000-04-20  
PRIOR APPLICATION NUMBER: US 60/193, 483  
PRIOR FILING DATE: 2000-03-29  
PRIOR APPLICATION NUMBER: US 60/185, 218  
PRIOR FILING DATE: 2000-02-24  
PRIOR APPLICATION NUMBER: US 60/167, 363  
PRIOR FILING DATE: 1999-11-23  
PRIOR APPLICATION NUMBER: US 60/156, 358  
PRIOR FILING DATE: 1999-09-28  
PRIOR APPLICATION NUMBER: US 60/146, 002  
PRIOR FILING DATE: 1999-08-09  
NUMBER OF SEQ ID NOS: 325720  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 287050  
LENGTH: 600  
TYPE: DNA  
ORGANISM: Human  
US-10-027-632-287050  
Query Match 53.0%; Score 75.8; DB 13; Length 600;  
Best local Similarity 78.5%; Pred. No. 2.1e-15;

Matches 117; Conservative 1; Mismatches 23; Indels 8; Gaps 2;

QY 1 GATTTGTTGCTGCTCTCTGGAGCAGCCTTTGCTATGCCCTGAGTAAATACCCCT 60  
|||  
DB 245 GATTTATTTGCTGCTCTCTGGAGCAGCCTTTGCTATGCCCTGAGTAAATACCCCT 186  
|||

QY 61 -----GCATATATTCATTTCAAGCTTGAATATAAGTGTGCCCAACAGTTGT 114  
|||  
DB 185 TGCATAGTCAGTGTCCATTTCAAACTTGGACATATAAATCTGCTC--ATAGTTGT 128  
|||

QY 115 AACTTTAGGTTTAAAGACAGTACAGATC 143  
|||  
DB 127 GAAATTAGGTTTAAACAGTATGAGATC 99  
|||

RESULT 2  
US-10-027-632-287050/c  
; Sequence 287050, Application US/10027632  
; Publication No. US20030204075A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; PRIOR FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 287050  
; LENGTH: 600  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-287050

Query Match 53.0%; Score 75.8; DB 15; Length 600;  
Best Local Similarity 78.5%; Pred. No. 2.1e-15;  
Matches 117; Conservative 1; Mismatches 23; Indels 8; Gaps 2;

QY 1 GATTTGTTGCTGCTCTCTGGAGCAGCCTTTGCTATGCCCTGAGTAAATACCCCT 60  
|||  
DB 245 GATTTATTTGCTGCTCTCTGGAGCAGCCTTTGCTATGCCCTGAGTAAATACCCCT 186  
|||

QY 61 -----GCATATATTCATTTCAAGCTTGAATATAAGTGTGCCCAACAGTTGT 114  
|||  
DB 185 TGCATAGTCAGTGTCCATTTCAAACTTGGACATATAAATCTGCTC--ATAGTTGT 128  
|||

QY 115 AACTTTAGGTTTAAAGACAGTACAGATC 143  
|||  
DB 127 GAAATTAGGTTTAAACAGTATGAGATC 99  
|||

RESULT 3  
US-10-027-632-287049/c  
; Sequence 287049, Application US/10027632  
; Publication No. US20020198371A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; PRIOR FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 287049  
; LENGTH: 600  
; TYPE: DNA  
; ORGANISM: Human

FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 287049  
; LENGTH: 600  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-287049

Query Match 52.2%; Score 74.6; DB 13; Length 600;  
Best Local Similarity 78.5%; Pred. No. 5.4e-15;  
Matches 117; Conservative 0; Mismatches 24; Indels 8; Gaps 2;

QY 1 GATTTGTTGCTGCTCTCTGGAGCAGCCTTTGCTATGCCCTGAGTAAATACCCCT 60  
|||  
DB 245 GATTTATTTGCTGCTCTCTGGAGCAGCCTTTGCTATGCCCTGAGTAAATACCCCT 186  
|||

QY 61 -----GCATATATTCATTTCAAGCTTGAATATAAGTGTGCCCAACAGTTGT 114  
|||  
DB 185 TGCATAGTCAGTGTCCATTTCAAACTTGGACATATAAATCTGCTC--ATAGTTGT 128  
|||

QY 115 AACTTTAGGTTTAAAGACAGTACAGATC 143  
|||  
DB 127 GAAATTAGGTTTAAACAGTATGAGATC 99  
|||

RESULT 4  
US-10-027-632-287049/c  
; Sequence 287049, Application US/10027632  
; Publication No. US20030204075A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 287049  
; LENGTH: 600  
; TYPE: DNA  
; ORGANISM: Human



```

; OPERATING SYSTEM: Microsoft Windows 95
; SOFTWARE: Wordperfect for windows 6.0,
; ASCII (DOS) TEXT format
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/455,150
; FILING DATE: 04-Jun-2003
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/324,709A
; FILING DATE: 03-June-1999
; ATTORNEY/AGENT INFORMATION:
; NAME: David J. Weitz
; REGISTRATION NUMBER: 38,362
; REFERENCE/DOCKET NUMBER: 16842-758
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 493-6811
; TELEFAX: (650) 493-9300
;
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 500 nucleotides
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
; SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-10-455-150-3
;
Query Match      26.0%; Score 37.2; DB 16; Length 500;
Best Local Similarity 92.9%; Pred. No. 0.032; 3; Indels 0; Gaps 0;
Matches 39; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 GATTTGTTTGCTGCTGCTGGAGAGAGCCTTTGCTATGCC 42
DB      80 GATTTATTTGCTGCTGCTGGAGAGAGCCTTTGCCATGCC 121

RESULT 9
US-10-104-774-1
; Sequence 1, Application US/10104774
; Publication No. US20020164630A1
; GENERAL INFORMATION:
; APPLICANT: Perkin-Elmer Corporation, Applied Biosystems Division
; TITLE OF INVENTION: DETERMINATION OF GENOTYPE OF AMPLIFICATION PRODUCTS AT MULTIPLE A
; FILE REFERENCE: 16842-782
; CURRENT APPLICATION NUMBER: US/10/104,774
; CURRENT FILING DATE: 2002-03-21
; PRIOR APPLICATION NUMBER: US 09/018,595
; PRIOR FILING DATE: 1998-02-04
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 793
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-104-774-1
;
Query Match      26.0%; Score 37.2; DB 13; Length 793;
Best Local Similarity 92.9%; Pred. No. 0.038; 3; Indels 0; Gaps 0;
Matches 39; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 GATTTGTTTGCTGCTGCTGGAGAGAGCCTTTGCTATGCC 42
DB      80 GATTTATTTGCTGCTGCTGGAGAGAGCCTTTGCCATGCC 121

RESULT 10
US-10-455-150-1
; Sequence 1, Application US/10455150
; Publication No. US20040053302A1
; GENERAL INFORMATION:
; APPLICANT: Perkin-Elmer Corporation,
; Applied Biosystems Division
; TITLE OF INVENTION: FLUORESCENCE GENOTYPING AT MULTIPLE ALLELIC
```

```

; SITES
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David J. Weitz,
; Wilson Soncini Goodrich & Rosati
; STREET: 650 Page Mill Road
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1050
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch diskette
; COMPUTER: IBM compatible
; OPERATING SYSTEM: Microsoft Windows 95
; SOFTWARE: Wordperfect for windows 6.0,
; ASCII (DOS) TEXT format
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/455,150
; FILING DATE: 04-Jun-2003
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/324,709A
; FILING DATE: 03-June-1999
; ATTORNEY/AGENT INFORMATION:
; NAME: David J. Weitz
; REGISTRATION NUMBER: 38,362
; REFERENCE/DOCKET NUMBER: 16842-758
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 493-9300
; TELEFAX: (650) 493-6811
;
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 793 nucleotides
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-10-455-150-1
;
Query Match      26.0%; Score 37.2; DB 16; Length 793;
Best Local Similarity 92.9%; Pred. No. 0.038; 3; Indels 0; Gaps 0;
Matches 39; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 GATTTGTTTGCTGCTGCTGGAGAGAGCCTTTGCTATGCC 42
DB      80 GATTTATTTGCTGCTGCTGGAGAGAGCCTTTGCCATGCC 121

RESULT 11
US-10-755-889-605
; Sequence 605, Application US/10755889
; Publication No. US20040171823A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ASSOCIATED WITH THE NF-KB
; FILE REFERENCE: D0284 NP
; CURRENT APPLICATION NUMBER: US/10/755,889
; CURRENT FILING DATE: 2004-01-13
; PRIOR APPLICATION NUMBER: U.S. 60/440,068
; PRIOR FILING DATE: 2003-01-14
; PRIOR APPLICATION NUMBER: U.S. 60/469,757
; PRIOR FILING DATE: 2003-05-12
; NUMBER OF SEQ ID NOS: 823
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 605
; LENGTH: 793
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-755-889-605
;
Query Match      26.0%; Score 37.2; DB 17; Length 793;
Best Local Similarity 92.9%; Pred. No. 0.038; 3; Indels 0; Gaps 0;
```



; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FASTSEQ for Windows Version 4.0  
; SEQ ID NO 218969  
; LENGTH: 680  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-218969

Query Match 22.8%; Score 32.6; DB 13; Length 680;  
Best Local Similarity 60.9%; Pred. No. 1.4;  
Matches 53; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 14 TGGCTCTGGGAGAGCCTTGCTATGCGCGTGAATTAATACCCCTGCATATATTCAA 73  
DB 676 TGGGGCTGGGAGAGCCTTGCTATGCGCGTGAATTAATACCCCTGCATATATTCAA 73  
QY 74 TTTCACAAAGCTTGGAATTAAGTCTG 100  
DB 616 TGTGACAAAGCATGATATAATAATATG 590

Search completed: February 9, 2005, 06:51:08  
Job time : 196.692 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

# OM nucleic - nucleic search, using sw model

Run on: February 8, 2005, 12:58:25 ; Search time 1703.26 Seconds  
(without alignments)  
3059.349 Million cell updates/sec

Title: US-10-754-437-10

Perfect score: 143  
1 gattctgttcgtcgtcctcc.....gtttaagacagtaagaatc 143

Sequence: 1 gattctgttcgtcgtcctcc.....gtttaagacagtaagaatc 143

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Database : Listing first 45 summaries

EST:\*  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_est9:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	41.8	29.2	303	W33932	W33932 mb54d02.r1
2	41.8	29.2	430	W40649	W40649 mc43h12.r1
3	41.8	29.2	536	W36345	W36345 mb72c12.r1
4	41.8	29.2	585	AY419441	AY419441 Mus muscu
5	41.8	29.2	645	CD773419	CD773419 AGENCOURT
6	41.8	29.2	707	CB056709	CB056709 NISC_j119
7	41.8	29.2	812	AK029358	AK029358 Mus muscu
8	41.8	29.2	843	CB588212	CB588212 AGENCOURT
9	41.8	29.2	859	CB588525	CB588525 AGENCOURT
10	41.8	29.2	865	CB589177	CB589177 AGENCOURT
11	41.8	29.2	886	CB587051	CB587051 AGENCOURT
12	41.8	29.2	889	CB589251	CB589251 AGENCOURT
13	41.8	29.2	891	CB574837	CB574837 AGENCOURT
14	41.8	29.2	913	CB590451	CB590451 AGENCOURT
15	41.8	29.2	919	CB587332	CB587332 AGENCOURT
16	41.8	29.2	928	CB590111	CB590111 AGENCOURT
17	41.8	29.2	945	BB614068	BB614068 BB614068
18	41.8	29.2	990	W12906	W12906 ma89p03.r1
19	41.8	29.2	1006	W29475	W29475 md99f11.r1
20	41.8	29.2	1020	W08102	W08102 mb40g08.r1
21	40.2	28.1	217	R46913	R46913 Y140 Rat in
22	40.2	28.1	223	BB571643	BB571643 BB571643
23	40.2	28.1	449	R47024	R47024 Y350 Rat in
24	40.2	28.1	1031	CB575508	CB575508 AGENCOURT

25	39.2	27.4	373	7	R47135	R47135 Y722 Rat in
26	39.2	27.4	395	7	R47030	R47030 Y359 Rat in
27	39.2	27.4	556	7	R46934	R46934 Y169 Rat in
28	38.6	27.0	257	7	R47100	R47100 Y601 Rat in
29	38.2	26.7	303	7	R46903	R46903 Y124 Rat in
30	38.2	26.7	318	7	R47078	R47078 Y534 Rat in
31	38.2	26.7	467	7	R46947	R46947 Y186 Rat in
32	37.6	26.3	223	7	R47143	R47143 Y79 Rat inc
33	37.6	26.3	720	2	BB368600	BB368600 BB368600
34	37.6	26.3	1444	3	AK048251	AK048251 Mus muscu
35	37.2	26.0	429	6	CB473314	CB473314 sm68 E03.
36	37.2	26.0	570	9	AY419439	AY419439 Homo sapi
37	37.2	26.0	570	9	AY419440	AY419440 Pan trogl
38	37	25.9	395	7	R46955	R46955 Y195 Rat in
39	36.6	25.6	478	7	R46933	R46933 Y166 Rat in
40	36.2	25.3	273	7	R47002	R47002 Y3 Rat inc1
41	36	25.2	1005	9	CL040441	CL040441 CH216-49N
42	35.6	24.9	286	7	R46887	R46887 Y106 Rat in
43	35.6	24.9	783	4	BG198114	BG198114 R5717499
44	35.2	24.6	513	8	AZ725508	AZ725508 R5717499
45	35	24.5	355	7	R46940	R46940 Y174 Rat in

## ALIGNMENTS

RESULT 1  
W33932  
LOCUS  
DEFINITION  
mb54d02.r1 Soares mouse p3NMF19.5 Mus musculus cDNA clone  
IMAGE:333219 5' similar to gb:M1005 Mouse amelogenin (MOUSE);,  
mRNA sequence.  
W33932  
W33932.1 GI:1315837  
VERSION  
W33932  
KEYWORDS  
EST.  
SOURCE  
Mus musculus (house mouse)  
ORGANISM  
Mus musculus  
REFERENCE  
1 (bases 1 to 303)  
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
Giesel,S., Kucaba,T., Lacey,M., Le,M., Martin,J., Morris,M.,  
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
Waterston,R.  
TITLE  
The WashU-HMI Mouse EST Project  
JOURNAL  
Unpublished (1996)  
COMMENT  
Contact: Marra M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.edu  
This clone is available royalty-free through LML; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:214619  
Seq primer: mob.REGA+ET  
High quality sequence stop: 185.  
Location/Qualifiers  
1..303  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:333219"  
/dev\_stage="19.5 dpc total fetus"  
/lab\_host="PH10B (ampicillin resistant)"  
/clone\_lib="Soares mouse p3NMF19.5"  
/note="Vector: pTR73D (Pharmacia) with a modified  
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA  
was primed with a Not I - oligo(dT) primer [5',  
TGTTCCATCTGAGTGGAGCGCGCATTTTCTTTTCTTTT 3']  
double-stranded cDNA was size selected, ligated to Eco RI

## ORIGIN

adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M. Fatima Bonaldo. RNA was kindly provided by Dr. Minoru Ko (Wayne State University)."

## Query Match

Best Local Similarity 29.2%; Score 41.8; DB 7; Length 303;

Matches 43; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GATTGTGTTGCTGCTCTCTGGAGAGAGCTTTGCTATGCCCGT 45

Db 81 GATTGTGTTGCTGCTCTCTGGAGAGAGCTTTGCTATGCCCGT 125

## RESULT 2

## LOCUS

W40649 430 bp mRNA linear EST 11-SEP-1996

DEFINITION mc43h12.r1 Soares mouse p3NMF19.5 Mus musculus cDNA clone

IMAGE:351335 5' similar to gb:M10095 Mouse amelogenin (MOUSE);,

mRNA sequence.

W40649.1 GI:1324982

EST.

Mus musculus (house mouse)

## SOURCE

Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 430)

Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,

Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,

Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,

Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and

Waterston,R.

The WashU-HMI Mouse EST Project

Unpublished (1996)

Contact: Marra M/Mouse EST Project

WashU-HMI Mouse EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

MG1:223335

Seq primer: EPrimer

High quality sequence stop: 323.

Location/Qualifiers

1. 430

/organism="Mus musculus"

/mol\_type="mRNA"

/db\_xref="taxon:10090"

/clone="IMAGE:351335"

/dev\_stage="19.5 dpc total fetus"

/lab\_host="DH10B (ampicillin resistant)"

/clone\_1ib="Soares mouse p3NMF19.5"

/note="Vector: pT73D (Pharmacia) with a modified

polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA

was primed with a Not I - oligo(dT) primer [5'

TGTTACCAATCTGAGTGGAGCGGCGGATTTTCTATGCCCGT 3'],

double-stranded cDNA was size selected, ligated to Eco RI

adapters (Pharmacia), digested with Not I and cloned into

the Not I and Eco RI sites of a modified pT73 vector

(Pharmacia). Library went through one round of

normalization to a Cot = 5. Library constructed by Bento

Soares and M. Fatima Bonaldo. RNA was kindly provided by

Dr. Minoru Ko (Wayne State University)."

ORIGIN

Query Match

29.2%; Score 41.8; DB 7; Length 430;

Best Local Similarity 95.6%; Pred. No. 0.012;

Matches 43; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GATTGTGTTGCTGCTCTCTGGAGAGAGCTTTGCTATGCCCGT 45

Db 69 GATTGTGTTGCTGCTCTCTGGAGAGAGCTTTGCTATGCCCGT 113

## RESULT 3

## LOCUS

W36345 536 bp mRNA linear EST 11-SEP-1996

DEFINITION mb72c12.r1 Soares mouse p3NMF19.5 Mus musculus cDNA clone

IMAGE:334966 5' similar to gb:M10095 Mouse amelogenin (MOUSE);,

mRNA sequence.

W36345.1 GI:1318120

## EST.

Mus musculus (house mouse)

## SOURCE

Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 536)

Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,

Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,

Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,

Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and

Waterston,R.

The WashU-HMI Mouse EST Project

Unpublished (1996)

Contact: Marra M/Mouse EST Project

WashU-HMI Mouse EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

MG1:216366

Seq primer: EPrimer

High quality sequence stop: 359.

Location/Qualifiers

1. 536

/organism="Mus musculus"

/mol\_type="mRNA"

/db\_xref="taxon:10090"

/clone="IMAGE:334966"

/dev\_stage="19.5 dpc total fetus"

/lab\_host="DH10B (ampicillin resistant)"

/clone\_1ib="Soares mouse p3NMF19.5"

/note="Vector: pT73D (Pharmacia) with a modified

polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA

was primed with a Not I - oligo(dT) primer [5'

TGTTACCAATCTGAGTGGAGCGGCGGATTTTCTATGCCCGT 3'],

double-stranded cDNA was size selected, ligated to Eco RI

adapters (Pharmacia), digested with Not I and cloned into

the Not I and Eco RI sites of a modified pT73 vector

(Pharmacia). Library went through one round of

normalization to a Cot = 5. Library constructed by Bento

Soares and M. Fatima Bonaldo. RNA was kindly provided by

Dr. Minoru Ko (Wayne State University)."

ORIGIN

Query Match

29.2%; Score 41.8; DB 7; Length 536;

Matches 43; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GATTGTGTTGCTGCTCTCTGGAGAGAGCTTTGCTATGCCCGT 45

Db 64 GATTGTGTTGCTGCTCTCTGGAGAGAGCTTTGCTATGCCCGT 108

## RESULT 4

AY419441 585 bp DNA linear GSS 17-DEC-2003  
LOCUS  
DEFINITION Mus musculus AMELX gene, VIRUTAL TRANSCRIPT, partial sequence,  
AY419441  
ACCESSION  
VERSION AY419441.1 GI:39775398  
KEYWORDS  
SOURCE  
ORGANISM Mus musculus (house mouse)  
REFERENCE  
AUTHORS Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 585)  
Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejarawal, A.,  
Todd, M.A., Tanenbaum, D.M., Civeille, D.R., Lu, F., Murphy, B.,  
Fertiera, S., Wang, G., Zheng, X.H., White, T.J., Smitsky, J.J.,  
Adams, M.D. and Cargill, M.  
Inferring nonneutral evolution from human-chimp-mouse orthologous  
gene trios  
Science 302 (5652), 1960-1963 (2003)  
14671302  
TITLE  
JOURNAL  
PUBMED  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
FEATURES  
source  
1.585  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
<1..>585  
/gene="AMELX"  
/locus\_tag="HCM6891"

ORIGIN  
Query Match 29.2%; Score 41.8; DB 9; Length 585;  
Best Local Similarity 95.6%; Pred. No. 0.013; Indels 0; Gaps 0;  
Matches 43; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 GATTTTGTTCCTGCTCCTCGAGACAGCTTTGCTATGCCCT 45  
12 GATTTTGTTCCTGCTCCTCGAGACAGCTTTGCTATGCCCT 56

RESULT 5  
CD773419 645 bp mRNA linear EST 02-JUL-2003  
LOCUS  
DEFINITION AGENCOURT 14713373 NIH MGC 190 Mus musculus cDNA clone  
IMAGE:30501267 5', mRNA sequence.  
CD773419  
ACCESSION  
VERSION CD773419.1 GI:32431921  
KEYWORDS  
SOURCE  
ORGANISM Mus musculus (house mouse)  
REFERENCE  
AUTHORS Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 645)  
NIH-MGC http://mgc.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Daniela S. Gerhard, Ph.D.  
Office of Cancer Genomics  
National Cancer Institute / NIH  
Bldg. 31 Km10A07 Bethesda, MD 20892  
Email: cgabs-remail.nih.gov  
Tissue Procurement: Yoshihiko Yamada, Takashi Nakamura, NIDCR  
CDNA Library Preparation: CLONTECH Laboratories, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
http://image.llnl.gov  
Plate: NDCM223 row: k column: 04  
High quality sequence stop: 611.  
location/Qualifiers  
1.645  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:30501267"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_1b="NIH MGC 190"  
/note="Organ: Pooled - Molar; Vector: pDNR-LIB; Site\_1:  
Site1 (ggccatctggcc); Site 2: Site1 (ggccgctggcc);  
Non-normalized full-length enriched library 5' and 3'  
adaptors were used in cloning as follows: 5' adaptor  
sequence: 5'-CACGGCCATTATGACC-3' and 3' adaptor sequence:  
5'-ATTCTAGAGCCGAGCGCCGACATG-dt(30)-BN-3' (where B = A,  
C, or G and N = A, C, G, or T). Average insert size 1.71  
kb (range 0.5-3.0 kb). 15/15 colonies contained inserts by  
PCR. This library was enriched for full-length clones and  
was constructed by Clontech Laboratories (Palo Alto, CA)  
Corp."

ORIGIN  
Query Match 29.2%; Score 41.8; DB 6; Length 645;  
Best Local Similarity 95.6%; Pred. No. 0.013; Indels 0; Gaps 0;  
Matches 43; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 GATTTTGTTCCTGCTCCTCGAGACAGCTTTGCTATGCCCT 45  
84 GATTTTGTTCCTGCTCCTCGAGACAGCTTTGCTATGCCCT 128

RESULT 6  
CB056709 707 bp mRNA linear EST 17-JAN-2003  
LOCUS  
DEFINITION NISC\_j11905.w1 Soares NMBP13-15 Mus musculus cDNA clone  
IMAGE:4848584 5', mRNA sequence.  
CB056709  
ACCESSION  
VERSION CB056709.1 GI:27794996  
KEYWORDS  
SOURCE  
ORGANISM Mus musculus (house mouse)  
REFERENCE  
AUTHORS Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 707)  
NCI-CCAP http://www.ncbi.nlm.nih.gov/ccgap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: cgabs-remail.nih.gov  
CDNA Library Preparation: The I.M.A.G.E. Consortium/LNL  
DNA Sequencing by: National Institutes of Health Intramural  
Sequencing Center (NISC)  
Clone distribution: NCI-CCAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
infoimage.llnl.gov  
Plate: LLM10779 row: N column: 9  
Seq primer: T7 primer.  
location/Qualifiers  
1.707  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:4848584"  
/tissue\_type="pituitary gland"  
/dev\_stage="juvenile, 13-15 days"  
/lab\_host="DH10B (phage-resistant)"

/clone.lib="Soares NMBP13-15"  
 /note="Organ: brain; Vector: pT73D-Pac1; Site 1: NotI;  
 Site 2: EcoRI, 1st strand cDNA was primed with a NotI-  
 oligo(dT) primer  
 5'-ACTGGAAGAGTTCGCGCGCGCTGTACCAAGTGTGTGTGTGTGTGT-3';  
 double-stranded cDNA was ligated to EcoRI adaptors  
 5'-ATTGGGACAGAG-3' and 5'-CCTGCGCG-3' (Pharmacia),  
 digested with NotI and cloned into the NotI and EcoRI  
 sites of the pT73D-Pac1 vector. Library went through one  
 round of normalization, and was constructed in the  
 laboratory of M. Bento Soares (University of Iowa)."

## ORIGIN

Query Match 29.2%; Score 41.8; DB 6; Length 707;  
 Best Local Similarity 95.6%; Pred. No. 0.013; Indels 0; Gaps 0;  
 Matches 43; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GATTTGTTGCTGCTGCTGCGAGCAGCCTTGTATGCCCGT 45  
 |||||||  
 Db 58 GATTTGTTGCTGCTGCTGCGAGCAGCCTTGTATGCCCGT 102

RESULT 7 AK029358 812 bp mRNA linear HTC 03-APR-2004  
 LOCUS AK029358 Mus musculus 0 day neonate head cDNA, RIKEN full-length enriched  
 DEFINITION AK029358 library, clone:4833404E21 product:amelogenin, full insert sequence.  
 ACCESSION AK029358  
 VERSION AK029358.1 GI:26325321  
 KEYWORDS HTC; CAP trapper.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
 AUTHORS Carninci, P. and Hayashizaki, Y.  
 TITLE High-efficiency full-length cDNA cloning  
 JOURNAL Meth. Enzymol. 303, 19-44 (1999)  
 MEDLINE 99279253  
 PUBMED 10349636  
 REFERENCE  
 AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
 Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
 TITLE Normalization and subtraction of cap-trapper-selected cDNAs to  
 JOURNAL prepare full-length cDNA libraries for rapid discovery of new genes  
 MEDLINE Genome Res. 10 (10), 1617-1630 (2000)  
 PUBMED 20499374  
 REFERENCE  
 AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagao, S., Sasaki, N., Carninci, P.,  
 Konno, H., Akiyama, J., Nishi, K., Katsunari, T., Tashiro, H., Itoh, M.,  
 Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,  
 Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,  
 Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M.,  
 Yonekura, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,  
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
 TITLE RIKEN integrated sequence analysis (RISA) system-384-format  
 JOURNAL sequencing pipeline with 384 multicapillary sequencer  
 MEDLINE Genome Res. 10 (11), 1757-1771 (2000)  
 PUBMED 20530913  
 REFERENCE  
 AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the  
 TITLE Functional annotation of a full-length mouse cDNA collection  
 JOURNAL Nature 409, 685-690 (2001)  
 REFERENCE  
 AUTHORS The PANTOM Consortium and the RIKEN Genome Exploration Research  
 TITLE Group Phase I & II Team.  
 JOURNAL Analysis of the mouse transcriptome based on functional annotation  
 REFERENCE of 60,770 full-length cDNAs  
 JOURNAL Nature 420, 563-573 (2002)  
 REFERENCE 6 (base 1 to 812)

## AUTHORS

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,  
 Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,  
 Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T.,  
 Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,  
 Kato, H., Kawai, Y., Kojima, Y., Konno, S., Konno, H., Kouda, M.,  
 Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,  
 Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,  
 Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,  
 Sano, H., Sasaki, D., Shibata, K., Shingawa, A., Shiraki, T.,  
 Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,  
 Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,  
 Muramatsu, M. and Hayashizaki, Y.

## TITLE

## JOURNAL

## COMMENT

Submitted (16-Jun-2001) Yoshinori Hayashizaki, The Institute of  
 Physical and Chemical Research (RIKEN), Laboratory for Genome  
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
 RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,  
 Kanagawa 230-0045, Japan (E-mail: genome-res@gscc.riken.jp,  
 URL: http://genome-gsc.riken.jp/, Tel: 81-45-503-9222,  
 Fax: 81-45-503-9216)  
 CDS  
 CDNA library was prepared and sequenced in Mouse Genome  
 Encyclopedia Project of Genome Exploration Research Group in Riken  
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
 Division of Experimental Animal Research in Riken contributed to  
 prepare mouse tissues.  
 Please visit our web site for further details.  
 URL: http://genome-gsc.riken.jp/  
 URL: http://fantom.gsc.riken.jp/  
 Location/Qualifiers

## FEATURES

## source

1..812  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="PANTOM:DB:4833404E21"  
 /db\_xref="taxon:10090"  
 /clone="4833404E21"  
 /tissue\_type="head"  
 /clone\_lib="RIKEN full-length enriched mouse cDNA library"  
 /dev\_stage="0 day neonate"  
 73..663  
 /note="unnamed protein product; amelogenin (MGD) MG1:88005,  
 GB|D17168, evidence: BLASTN, 99%, match=799)  
 putative"  
 /codon\_start=1  
 /protein\_id="BAC26415.1"  
 /db\_xref="GI:26325322"  
 /translation="MTWILFACLLGAAPMPHPESPFGYINLSFVLTPLKRYOS  
 MTRQYPSYGYEPMGWLHQIIPVLSQHPSPHTLDQHNHLPVPAQAPVAPQPM  
 PVPGHSMTPYQNHQIPNPSPAQPFQCPQPPQAIIPQSHQPMQFQSPHMQPLAPQ  
 PPLPLPFSMOPLSPLPLPLPLPAMPATDKTGREVD"  
 793..798  
 /note="putative"  
 812  
 /note="putative"

## CDS

polyA\_signal  
 polyA\_site

## ORIGIN

Query Match 29.2%; Score 41.8; DB 3; Length 812;  
 Best Local Similarity 95.6%; Pred. No. 0.014; Indels 0; Gaps 0;  
 Matches 43; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GATTTGTTGCTGCTGCTGCGAGCAGCCTTGTATGCCCGT 45  
 |||||||  
 Db 84 GATTTGTTGCTGCTGCTGCGAGCAGCCTTGTATGCCCGT 128

RESULT 8 CB588212 843 bp mRNA linear EST 03-APR-2003  
 LOCUS CB588212  
 DEFINITION AGENCOURT 12771359 NIH MGC 136 Mus musculus cDNA clone  
 ACCESSION CB588212  
 VERSION CB588212.1 GI:29506068  
 KEYWORDS EST.

SOURCE	Mus musculus (house mouse)
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE	1 (bases 1 to 843)
AUTHORS	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished (1999)
COMMENT	Contact: Robert Strausberg, Ph.D. Email: <a href="mailto:cgapbs-remail.nih.gov">cgapbs-remail.nih.gov</a> Tissue Procurement: Dr. David Rowe cDNA Library Preparation: Invitrogen Corp DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: <a href="http://image.llnl.gov">http://image.llnl.gov</a> Plate: NDAM338 row: 9 column: 08 High quality sequence accp: 636. Location/Qualifiers 1. 843 /organism="Mus musculus" /mol_type="mRNA" /db_xref="taxon:10090" /clone="IMAGE:30293239" /tissue_type="embryonic limb, maxilla and mandible" /lab_host="DH10B (phage-resistant)" /clone_1lib="NIH MGC 136" /note="Vector: pCMV-SPORT6.1; Site 1: EcoRV; Site 2: NotI, Normalized, full-length enriched library from pool of mouse embryonic limb, maxilla and mandible, embryonic day 17.5, 18.5 and newborn (mandible (5), 4 and 1 limb and jaw equivalents from respective days). Cloned directionally, o14go-dt primed (5'-GACTGATCTTGATCGCAGCGCGCCCT(7)15-3', size selected for the >1kb fragments, average insert size 1.2 kb. Normalization to Cot 7.5. Tissue contributed by David Rowe; library constructed by Reegen, Invitrogen Corp. Note: this is a NIH_MGC Library."
ORIGIN	
Query Match	29.2%; Score 41.8; DB 6; Length 843;
Best Local Similarity	95.6%; Pred. No. 0.014; Mismatches 2; Indels 0; Gaps 0;
Matches	43; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy	1 GATTTTGTTTGCCCTCCTCGGAGCAGCCTTTCCTATGCGCGGT 45       51 GATTTGTGCTGCTCCCTCGGAGCAGCCTTTCCTATGCGCGCT 95
Db	
RESULT 9	
LOCUS	CB588525
DEFINITION	CB588525 859 bp mRNA linear EST 03-APR-2003
VERSION	IMGC:30289751.5, mRNA sequence.
VERSION	CB588525
KEYWORDS	CB588525.1 GI:29506381
SOURCE	EST.
ORGANISM	Mus musculus (house mouse)
REFERENCE	Mus musculus
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT	Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: <a href="mailto:cgapbs-remail.nih.gov">cgapbs-remail.nih.gov</a> Tissue Procurement: Dr. David Rowe cDNA Library Preparation: Invitrogen Corp DNA Sequencing by: The I.M.A.G.E. Consortium (LNLN) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: 1 (bases 1 to 859) NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .

FEATURES	source
http://image.lnl.gov plate: NDAM329 row: e column: 24 High quality sequence stop: 763. Location/Qualifiers	1. .859 /organism="Mus musculus" /mol_type="mRNA" /db_xref="taxon:10090" /clone="IMAGE:30289751" /tissue_type="embryonic limb, maxilla and mandible" /lab_host="DH10B (phage-resistant)" /clone_lib="NIH MGC 136" /note="Vector: PCMV-SPORT6.1; Site 1: EcoRV; Site 2: NotI; Normalized, full-length enriched library from pool of mouse embryonic limb, maxilla and mandible, embryonic day 17.5, 18.5 and newborn (mandible (5, 4 and 1 limb and jaw equivalents from respective days). Cloned directionally, oligo-dT primed (5'-GACATCTTCATGATCGGACGCGCCCTT)15-3'. Size selected for the >1kb fragments, average insert size 1.2 kb. Normalization to Cot 7.5. Tissue contributed by David Rowe; library constructed by ResGen, Invitrogen Corp. Note: this is a NIH_MGC Library."
ORIGIN	
Query Match 29.2%; Score 41.8; DB 6; Length 859; Best Local Similarity 95.6%; Pred. No. 0.014; Matches 43; Conservative 0; Mismatches 2; Indels 0; Gaps 0;	
0y	1 GATTGTGTCCTGCGCTCGGAGAGACGCTTGATGCCCGT 45       70 GATTGTGTCCTGCGCTCGGAGAGACGCTTGATGCCCGCT 114
RESULT 10	
LOCUS CB589177	
DEFINITION AGENCOURT_12770446 NIH_MGC_136 Mus musculus cDNA clone	
ACCESSION IMAGE:30290534 5', mRNA sequence.	
VERSION CB589177.	
KEYWORDS CB589177.1 GI:29507033	
SOURCE EST.	
ORGANISM Mus musculus (house mouse)	
REFERENCE Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 865)	
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.	
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)	
JOURNAL Unpublished (1999)	
COMMENT Contact: Robert Strauberg, Ph.D. Email: cgabds-remail.nih.gov Tissue Procurement: Dr. David Rowe cDNA Library Preparation: Invitrogen Corp cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov plate: NDAM331 row: f column: 15 High quality sequence stop: 630. Location/Qualifiers	
FEATURES	
source	1. .865 /organism="Mus musculus" /mol_type="mRNA" /db_xref="taxon:10090" /clone="IMAGE:30290534" /tissue_type="embryonic limb, maxilla and mandible" /lab_host="DH10B (phage-resistant)" /clone_lib="NIH MGC 136" /note="Vector: PCMV-SPORT6.1; Site 1: EcoRV; Site 2: NotI; Normalized, full-length enriched library from pool of mouse embryonic limb, maxilla and mandible, embryonic day 17.5, 18.5 and newborn (mandible (5, 4 and 1 limb and jaw



TITLE National Incestutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-remail.nih.gov  
Tissue Procurement: Dr. David Rowe  
CDNA Library Preparation: Invitrogen Corp  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:  
http://image.llnl.gov  
Plate: NDA342 row: d column: 24  
High quality sequence start: 19  
High quality sequence stop: 658.  
Location/Qualifiers

## FEATURES

1. .919  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:30294719"  
/tissue\_type="embryonic limb, maxilla and mandible"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_1lb="NIH\_MGC\_136"  
/note="Vector: PCMV-SPORT6.1; Site 1: EcoRV; Site 2: NotI; Normalized, full-length enriched library from pool of mouse embryonic limb, maxilla and mandible, embryonic day 17.5, 18.5 and newborn (mandible (5, 4 and 1 limb and jaw equivalents from respective days). Cloned directionally, oligo-dt primed (5'-GACTAGTTTACATCGACGCCGCC(T)15-3'. Size selected for the >1kb fragments, average insert size 1.2 kb. Normalization to Cot 7.5. Tissue contributed by David Rowe; library constructed by Resgen, Invitrogen Corp. Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 29.2%; Score 41.8; DB 6; Length 891;  
Best Local Similarity 95.6%; Pred. No. 0.014;  
Matches 43; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GATTTGTTGCTGCTCTCGAGACGCTTTGCTATGCCCT 45  
|||||  
Db 90 GATTTGTTGCTGCTCTCGAGACGCTTTGCTATGCCCT 134

RESULT 14 913 bp mRNA linear EST 03-APR-2003  
CBS90451  
LOCUS AGENCOURT\_12770290 NIH\_MGC\_136 Mus musculus CDNA clone  
DEFINITION IMAGE:30286944 5', mRNA sequence.  
ACCESSION CBS90451  
VERSION CBS90451.1 GI:29508307  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 913)  
NIH-MGC http://mgc.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-remail.nih.gov  
Tissue Procurement: Dr. David Rowe  
CDNA Library Preparation: Invitrogen Corp  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:  
http://image.llnl.gov  
Plate: NDA342 row: a column: 01  
High quality sequence start: 12  
High quality sequence stop: 640.  
Location/Qualifiers

## FEATURES

1. .919  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:30294719"  
/tissue\_type="embryonic limb, maxilla and mandible"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_1lb="NIH\_MGC\_136"  
/note="Vector: PCMV-SPORT6.1; Site 1: EcoRV; Site 2: NotI; Normalized, full-length enriched library from pool of mouse embryonic limb, maxilla and mandible, embryonic day 17.5, 18.5 and newborn (mandible (5, 4 and 1 limb and jaw equivalents from respective days). Cloned directionally, oligo-dt primed (5'-GACTAGTTTACATCGACGCCGCC(T)15-3'. Size selected for the >1kb fragments, average insert size 1.2 kb. Normalization to Cot 7.5. Tissue contributed by David Rowe; library constructed by Resgen, Invitrogen Corp. Note: this is a NIH\_MGC Library."

## source

1. .913  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:30286944"  
/tissue\_type="embryonic limb, maxilla and mandible"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_1lb="NIH\_MGC\_136"  
/note="Vector: PCMV-SPORT6.1; Site 1: EcoRV; Site 2: NotI; Normalized, full-length enriched library from pool of mouse embryonic limb, maxilla and mandible, embryonic day 17.5, 18.5 and newborn (mandible (5, 4 and 1 limb and jaw equivalents from respective days). Cloned directionally, oligo-dt primed (5'-GACTAGTTTACATCGACGCCGCC(T)15-3'. Size selected for the >1kb fragments, average insert size 1.2 kb. Normalization to Cot 7.5. Tissue contributed by David Rowe; library constructed by Resgen, Invitrogen Corp. Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 29.2%; Score 41.8; DB 6; Length 913;  
Best Local Similarity 95.6%; Pred. No. 0.014;  
Matches 43; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GATTTGTTGCTGCTCTCGAGACGCTTTGCTATGCCCT 45  
|||||  
Db 82 GATTTGTTGCTGCTCTCGAGACGCTTTGCTATGCCCT 126

RESULT 15 919 bp mRNA linear EST 03-APR-2003  
CBS87332  
LOCUS AGENCOURT\_12971606 NIH\_MGC\_136 Mus musculus CDNA clone  
DEFINITION IMAGE:30291971 5', mRNA sequence.  
ACCESSION CBS87332  
VERSION CBS87332.1 GI:29505188  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 919)  
NIH-MGC http://mgc.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-remail.nih.gov  
Tissue Procurement: Dr. David Rowe  
CDNA Library Preparation: Invitrogen Corp  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:  
http://image.llnl.gov  
Plate: NDA335 row: b column: 12  
High quality sequence start: 216  
High quality sequence stop: 325.  
Location/Qualifiers

## FEATURES

1. .919  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:30291971"  
/tissue\_type="embryonic limb, maxilla and mandible"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_1lb="NIH\_MGC\_136"  
/note="Vector: PCMV-SPORT6.1; Site 1: EcoRV; Site 2: NotI; Normalized, full-length enriched library from pool of mouse embryonic limb, maxilla and mandible, embryonic day 17.5, 18.5 and newborn (mandible (5, 4 and 1 limb and jaw equivalents from respective days). Cloned directionally, oligo-dt primed (5'-GACTAGTTTACATCGACGCCGCC(T)15-3'. Size selected for the >1kb fragments, average insert size 1.2 kb. Normalization to Cot 7.5. Tissue contributed by David Rowe; library constructed by Resgen, Invitrogen Corp. Note: this is a NIH\_MGC Library."









Db 2119 GATTTCTTGGCTGCTTTGGAGAGCATCCCTGTATGCTGTGATTAAG--CTA 2176

QY 61 TTTCATTTCAAGCTTGAATAAATCTCCCTCAGA 101

Db 2177 AGTCAATTTCAATGCTTGAAATAAATTCACACATA 2217

RESULT 4  
AB091781 6465 bp DNA linear PRI 02-MAY-2003  
LOCUS Pan troglodytes AMELX gene for amelogenin, partial cds.  
DEFINITION  
AB091781  
VERSION  
AB091781.1 GI:29126015  
KEYWORDS  
SOURCE  
ORGANISM  
Pan troglodytes (chimpanzee)  
Pan troglodytes  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Pan.

REFERENCE  
AUTHORS  
TITLE  
1 Iwase, M., Saita, Y., Hirai, Y., Hirai, H., Imai, H. and Takahata, N.  
From the Cover: The amelogenin loci span an ancient pseudautosomal  
boundary in diverse mammalian species  
JOURNAL  
MEDLINE  
22608569  
Proc. Natl. Acad. Sci. U.S.A. 100 (9), 5258-5263 (2003)

REFERENCE  
AUTHORS  
TITLE  
2 (bases 1 to 6465)  
Iwase, M., Saita, Y. and Takahata, N.  
Direct Submission  
Submitted (19-SEP-2002) Mineyo Iwase, Graduate University for  
Advanced Studies (Soken), Department of Biosystems Science;  
Shonan Kokuasimura, Hayama, Kanagawa 240-0193, Japan  
(E-mail: iwase@min@koryu.w01.soken.ac.jp, Tel: 81-468-58-1571,  
Fax: 81-468-58-1544)

FEATURES  
source  
Location/Qualifiers  
1. 6465  
/organism="Pan troglodytes"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9598"  
/sex="male"  
927. 6465  
/gene="AMELX"  
join(927. 982,2283. 2294)  
/gene="AMELX"  
join(2295. 2348,4284. 4331,5607. 5648,5740. 5781,  
6052. 6465)  
/gene="AMELX"  
/codon\_start=1  
/product="amelogenin"  
/protein\_id="BAC66101.1"  
/db\_xref="GI:29126015"  
/translation="MGTWILFACLLGAAPMPRPHPGPGYINFSYENSHQAINVD  
RTALVLPKMYQSIIRPYSYGEBPMGMLHQLIPVLSQHPHTTLQPHHILPV  
PAAQPVLPQPPMPLPGQHSMTPTQHNQPLPPAAQPYQPPQPPQPPQPPQPPV  
HPMQPLPQPPPLPMPFMPQPLPMLPDLTLEAMPSTDKR"

ORIGIN  
Query Match 37.5%; Score 54.8; DB 9; Length 6465;  
Best Local Similarity 77.8%; Pred. No. 0.00015;  
Matches 84; Conservative 0; Mismatches 12; Indels 12; Gaps 1;

QY 1 GATTTGTTGGCTGCTCTGGAGAGCATCCCTGTATGCTGTGATTAATTC--- 57  
|||||  
Db 2306 GATTTATTTGCTGCTCTGGAGAGCATCCCTGTATGCTGTGATTAATTC--- 57  
|||||

QY 58 -----CAATTTCAATTTCAAGCTTGAATAAATCTGCTGCC 96  
|||||

Db 2366 TGCAATAGTCAGTCCCAATTTCAAGCTTGAATAAATCTGCTGCC 2413  
|||||

RESULT 5  
AB091791 7425 bp DNA linear MAM 02-MAY-2003  
LOCUS Sus scrofa AMELX gene for amelogenin, partial cds.  
DEFINITION

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
AB091791 GI:29126034  
Sus scrofa (pig)  
Sus scrofa  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.

REFERENCE  
AUTHORS  
TITLE  
1 Iwase, M., Saita, Y., Hirai, Y., Hirai, H., Imai, H. and Takahata, N.  
From the Cover: The amelogenin loci span an ancient pseudautosomal  
boundary in diverse mammalian species  
JOURNAL  
MEDLINE  
22608569  
Proc. Natl. Acad. Sci. U.S.A. 100 (9), 5258-5263 (2003)

REFERENCE  
AUTHORS  
TITLE  
2 (bases 1 to 7425)  
Iwase, M., Saita, Y. and Takahata, N.  
Direct Submission  
Submitted (19-SEP-2002) Mineyo Iwase, Graduate University for  
Advanced Studies (Soken), Department of Biosystems Science;  
Shonan Kokuasimura, Hayama, Kanagawa 240-0193, Japan  
(E-mail: iwase@min@koryu.w01.soken.ac.jp, Tel: 81-468-58-1571,  
Fax: 81-468-58-1544)

FEATURES  
source  
Location/Qualifiers  
1. 7425  
/organism="Sus scrofa"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9823"  
/sex="male"  
1491. 7425  
/gene="AMELX"  
join(1491. 1548,3140. 3151)  
/gene="AMELX"  
join(3152. 3205,5316. 5363,6619. 6660,6717. 6761,  
7033. 7425)  
/gene="AMELX"  
/codon\_start=1  
/product="amelogenin"  
/protein\_id="BAC66110.1"  
/db\_xref="GI:29126035"  
/translation="MGTWILFACLLGAAPMPRPHPGPGYINFSYEDLYLAIRID  
RTALVLPKMYQNMRIHPYSYGEBPMGMLHQLIPVLSQHPHTTLQPHHILPV  
PAAQPVLPQPPMPLPGQHSMTPTQHNQPLPPAAQPYQPPQPPQPPQPPQPPV  
HPMQPLPQPPPLPMPFMSQSLPDLTLEAMPAT"

ORIGIN  
Query Match 37.5%; Score 54.8; DB 4; Length 7425;  
Best Local Similarity 75.0%; Pred. No. 0.00015;  
Matches 87; Conservative 0; Mismatches 17; Indels 12; Gaps 1;

QY 1 GATTTGTTGGCTGCTCTGGAGAGCATCCCTGTATGCTGTGATTAATTCGAA 60  
|||||

Db 3163 GATTTGTTGGCTGCTCTGGAGAGCATCCCTGTATGCTGTGATTAATTCGAA 60  
|||||

QY 61 TT-----TCAATTTCAAGCTTGAATAAATCTGCTGAGATT 104  
|||||

Db 3223 TTCTACGTACATGCTCAATGCTATACGCTTGAAATAAATCTACCCGAGTT 3278  
|||||

RESULT 6  
AB091794 5591 bp DNA linear MAM 02-MAY-2003  
LOCUS Equus caballus AMELX gene for amelogenin, partial cds.  
DEFINITION  
AB091794  
VERSION  
AB091794.1 GI:29126040  
KEYWORDS  
SOURCE  
ORGANISM  
Equus caballus (horse)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Perissodactyla; Equidae; Equus.

REFERENCE  
AUTHORS  
TITLE  
1 Iwase, M., Saita, Y., Hirai, Y., Hirai, H., Imai, H. and Takahata, N.  
From the Cover: The amelogenin loci span an ancient pseudautosomal  
boundary in diverse mammalian species

FEATURES	source	Location/Qualifiers
gene	130..5591	/organism="Equus caballus" /mol_type="genomic DNA" /db_xref="taxon:9796" /sex="male"
5'UTR	130..5591	/gene="AMELY" join(130..180,1483..1494) /gene="AMELY" join(1495..1548,3470..3517,4907..4915,5235..5591) /gene="AMELY"
CDS		/codon_start=1 /product="amelogenin" /protein_id="BAC66113.1" /db_xref="GI:29126041" /translation="MGWITFLACLVGTALAMPRLPHHGPAYINFSYEVLTYPGSGYEMSGMHHQIIPVLTOQNSHALOPFNHRIIPVPAQOPVVSQCPAIPLPGSHMIPQ HHQNPRLPPVQPPQPPQPPHPIQPPPMHPIQPPRLPQAPLPIPIPLAASGPHAS"
ORIGIN		
Query Match	37.3%; Score 54.4; DB 4; Length 5591;	
Beet Local Similarity	71.2%; Pred. No. 0.0002;	
Matches 104; Conservative 0; Mismatches 31; Indels 11; Gaps 2;		
QY	1 GATTGTGTCGCGCCCTCGTGGGAGCAGCCTCAGTATGCTGAGTAAATTTCCAA 60	
DB	1506 GATTGTGTCGCGCCCTGTTGGACAGCCATTGCTATGCCGTAAATTAAT--CAG 1553	
QY	61 TTTCCAAATTCACAGCTTGGAATATAATTCGCTCAGATTTTATTAGGTTAAAT 120	
DB	1564 CATCTATTTCACAGCTTGGAATGCAAAAATGCGCTCATTTGT-----GCTAAAT 1614	
QY	121 TTAAAGCTTTAAGACGTACCAAGAC 146	
DB	1615 TTTGGGGTTTAATCAATACCAAGATC 1640	
RESULT 7		
AB091783	AB091783	6442 bp DNA linear PRI 02-MAY-2003
LOCUS	Saimiri sciureus AMELX gene for amelogenin, partial cds.	
DEFINITION	AB091783	
ACCESSION	AB091783.1 GI:29126019	
VERSION		
KEYWORDS		
SOURCE	Saimiri sciureus (common squirrel monkey)	
ORGANISM	Saimiri sciureus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Cebinae; Saimiri.	
REFERENCE	1 Iwase,M., Satta,Y., Hirai,Y., Hirai,H., Imai,H. and Takahata,N. AUTHORS TITLE From the cover: The amelogenin loci span an ancient pseudautosomal boundary in diverse mammalian species JOURNAL Proc. Natl. Acad. Sci. U.S.A. 100 (9), 5258-5263 (2003)	
MEDLINE	2 (bases 1 to 6442)	
PUBMED	12672962	
REFERENCE	Iwase,M., Satta,Y. and Takahata,N. AUTHORS TITLE Direct Submision Submitted (19-SEP-2002) Mineyo Iwase, Graduate University for Advanced Studies (Sokenai), Department of Biosystems Science; Shonan Kofunakura, Hayama, Kanagawa 240-0193, Japan (E-mail:iwase@minokoryu01.soken.ac.jp, Tel:81-468-58-1571, Fax:81-468-58-1544)	

Shonan Kokusaimura, Hayama Kanagawa 240-0193, Japan  
[E-mail: iwase@minokoryu.w01.soken.ac.jp, Tel: 81-468-58-1571,  
Fax: 81-468-58-1544]  
Location/Qualifiers

FEATURES  
source  
1. 6442  
/organism="Saimiri sciureus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9521"  
/sex="male"  
936..6442  
/gene="AMELX"  
join(936..991,2298..2309)  
/gene="AMELX"  
join(2310..2363,4289..4336,5592..5633,5725..5769,  
6041..>6442)  
/gene="AMELX"  
/codon\_start=1  
/product="amelogenin"  
/protein\_id="BAC6103.1"  
/db\_xref="GI:29126020"  
/translation="MGWTLFACILGAFAPMP,PPHHGHPYINFSYENSHSOAIND  
RTGLVPLKYSQSMTRPATRYSGYEGMGLNHQITPLVSQQHPRNHTQPHNIPV  
VPAQDPVPOQPMVPVGGHSMIPQHNQENLPPRAQDPQVPOVOPQHPQMPQPP  
VHPQPLPQPLPPLPMPFMQPLPPLPDLPLAMPAT"

ORIGIN

Query Match 36.7%; Score 53.6; DB: 9; Length 6442;  
Best Local Similarity 74.6%; Pred. No. 0.0003;  
Matches 85; Conservative 0; Mismatches 19; Indels 10; Gaps 1;

QY 1 GATTTTGTTGGCCGCGCTCTCGGAGAGAGCCCTTCAGTATGCTGTGAGTAAATTTG--- 57  
|||||  
Db 2321 GATTTTGTTGCCGCTGCTCTCGGAGAGAGCTTTTGCCATGCTGTGAGTAAACCCCTG 2380  
|||||

QY 58 -----CAATTTCCAAATTTTCAAGGCTTGAAATAAATCTGCTCAGATT 104  
|||||  
Db 2381 CATAAGTCAGTGCATATTCATCAAGTATGCAATAAAGTCTGCCCATGTT 2434  
|||||

RESULT 8  
AB091786 8004 bp DNA linear PRI 02-MAY-2003  
LOCUS AB091786  
DEFINITION Lemur catca AMELY gene for amelogenin, partial cds.  
ACCESSION AB091786  
VERSION AB091786.1 GI:29126025  
KEYWORDS  
SOURCE  
ORGANISM  
Lemur catca (ring-tailed lemur)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Strepsirrhini; Lemuridae; Lemur.  
1  
Iwase, M., Satta, Y., Hiraai, Y., Hiraai, H., Imai, H. and Takahata, N.  
From the Cover: The amelogenin loci span an ancient pseudautosomal  
boundary in diverse mammalian species  
Proc. Natl. Acad. Sci. U.S.A. 100 (9), 5258-5263 (2003)  
22608569  
JOURNAL MEDLINE  
PUBMED 12672962  
2 (bases 1 to 8004)  
Iwase, M., Satta, Y. and Takahata, N.  
Direct Submission  
Submitted (19-SEP-2002) Mineyo Iwase, Graduate University for  
Advanced Studies(Sokenaai), Department of Biosystems Science;  
Shonan Kokusaimura, Hayama, Kanagawa 240-0193, Japan  
(E-mail: iwase@minokoryu.w01.soken.ac.jp, Tel: 81-468-58-1571,  
Fax: 81-468-58-1544)  
Location/Qualifiers  
1..8004  
/organism="Lemur catca"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9447"  
/sex="male"  
1382..6505  
/gene="AMELY"

5' UTR  
CDS  
join(1382..1437,3040..3051)  
/gene="AMELY"  
join(3052..3105,5192..5239,6461..>6505)  
/gene="AMELY"

/product="amelogenin"  
/protein\_id="BAC6106.1"  
/db\_xref="GI:29126026"  
/translation="MGTWILFACLLGAFAFMPLOPHRGHGYINFSYEVLTPLKWTOS  
MIRPP"

## ORIGIN

Query Match 36.7%; Score 53.6; DB 9; Length 8004;  
Best Local Similarity 78.6%; Pred. No. 0.00032;  
Matches 77; Conservative 0; Mismatches 19; Indels 2; Gaps 1;

QY 1 GATTTGTTGCTGCTGCTGAGAGAGCCTTCAGTATGCTGTGAGTAAATTTCCAA 60  
DB 3063 GATTTGTTGCTGCTGCTGAGAGAGCCTTTGTATGCTGTGAGTAAACATACCT 3122  
QY 61 TTTCACAA--TTTCACAGCTTGGAATAAATAATCTGCC 96  
DB 3123 TGCCTAAGTGTATAAAGCTTAAATATAAATCTGCC 3160

## RESULT 9

AB091787 5562 bp DNA linear PRI 02-MAY-2003  
LOCUS AB091787  
DEFINITION Otollemur garnettii AMELX gene for amelogenin, partial cds.  
ACCESSION AB091787  
VERSION AB091787.1 GI:29126027  
KEYWORDS Otollemur garnettii (small-eared galago)  
SOURCE Otollemur garnettii  
ORGANISM Otollemur garnettii  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Strepsirrhini; Galagonidae; Otollemur.

REFERENCE 1 Iwase, M., Saita, Y., Hirai, Y., Hirai, H., Imai, H. and Takahata, N.  
AUTHORS From the Cover: The amelogenin loci span an ancient pseudautosomal  
boundary in diverse mammalian species  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 100 (9), 5258-5263 (2003)  
MEDLINE 12672962  
PUBMED 12672962

REFERENCE 2 (bases 1 to 5562)  
AUTHORS Iwase, M., Saita, Y. and Takahata, N.  
TITLES Direct Submision  
JOURNAL Submitted (19-SEP-2002) Mrieyo Iwase, Graduate University for  
Advanced Studies (Soken), Department of Biosystems Science;  
Shonan Kokubaijima, Hayama, Kanagawa 240-0193, Japan  
(E-mail: iwase@nkcoryu01.soken.ac.jp, Tel: 81-468-58-1571,  
Fax: 81-468-58-1544)

## FEATURES

source  
1..5562  
Location/Qualifiers  
/organism="Otollemur garnettii"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:30611"  
/sex="male"  
267..5562  
/gene="AMELX"  
join(267..321,1401..1412)  
/gene="AMELX"  
join(1413..1466,3385..3432,4699..4740,4832..4876,  
5125..>5562)  
/gene="AMELX"  
/codon\_start=1  
/product="amelogenin"  
/protein\_id="BAC6107.1"  
/db\_xref="GI:29126028"  
/translation="MGTWILFACLLGAFAFMPLOPHRGHGYINFSYEVLTPLKWTOS  
MIRPP"

## gene

## 5' UTR

## CDS

join(1382..1437,3040..3051)  
/gene="AMELY"  
join(3052..3105,5192..5239,6461..>6505)  
/gene="AMELY"  
/product="amelogenin"  
/protein\_id="BAC6106.1"  
/db\_xref="GI:29126026"  
/translation="MGTWILFACLLGAFAFMPLOPHRGHGYINFSYEVLTPLKWTOS  
MIRPP"

## ORIGIN

Query Match 36.2%; Score 52.8; DB 9; Length 5562;  
Best Local Similarity 69.6%; Pred. No. 0.00053;  
Matches 110; Conservative 0; Mismatches 27; Indels 21; Gaps 2;

QY 1 GATTTGTTGCTGCTGCTGAGAGAGCCTTCAGTATGCTGTGAGTAAATTTTC--- 57  
DB 1424 GATTTGTTGCTGCTGCTGAGAGAGCTTTTGCATGCCGTGAGTAAACACCCCT 1483  
QY 58 -----CAATTCATTTTCAACAGCTTGAATAAATAATCTGCTCAGATTTTAT 108  
DB 1484 TGCATAAGTCAGTATCAATTTTCAATGCTTAAATAAATAATCTGCTCAACAGTT--- 1539  
QY 109 TTAGGCTTAATTTAAGCTTTAAGCAGTACAAAGATC 146  
DB 1540 -----GGTAACTTCAGGATTTAAACAGTATGATTC 1572

## RESULT 10

AY040206 8810 bp DNA linear PRI 10-APR-2002  
LOCUS AY040206  
DEFINITION Homo sapiens amelogenin precursor (AMELX) gene, complete cds.  
ACCESSION AY040206  
VERSION AY040206.1 GI:15028582  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 8810)  
AUTHORS Hart, P.S., Hart, T.C., Simer, J.P. and Wright, J.T.  
JOURNAL A nomenclature for X-linked amelogenesis imperfecta  
Arch. Oral Biol. 47 (4), 255-260 (2002)  
MEDLINE 21920287  
PUBMED 21920287

REFERENCE 2 (bases 1 to 8810)  
AUTHORS Hart, S., Hart, T.C. and Wright, T.  
TITLES Direct Submision  
JOURNAL Submitted (14-JUN-2001) Human Genetics, University of Pittsburgh,  
3550 Terrace St., 572A Scalle Hall, Pittsburgh, PA 15090, USA  
Location/Qualifiers  
1..8810  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="X"  
/map="Xp22.3-p22.1"  
1..8791  
/gene="AMELX"  
1..1416  
/gene="AMELX"  
1386..1391  
/gene="AMELX"  
join(1417..1472,2781..2846,4783..4830,6240..6281,  
6552..6977,8611..8791)  
/gene="AMELX"  
/product="amelogenin precursor"  
1417..1472  
/gene="AMELX"  
/number=1  
2781..2846  
/gene="AMELX"  
/number=2  
join(2793..2846,4783..4830,6240..6281,6552..6977,  
8611..8616)  
/gene="AMELX"  
/codon\_start=1  
/product="amelogenin precursor"  
/protein\_id="AAK77213.1"  
/db\_xref="GI:15028583"  
/translation="MGTWILFACLLGAFAFMPLOPHRGHGYINFSYEVLTPLKWTOS  
MIRPP"

## FEATURES

source  
1..8810  
Location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="X"  
/map="Xp22.3-p22.1"  
1..8791  
/gene="AMELX"  
1..1416  
/gene="AMELX"  
1386..1391  
/gene="AMELX"  
join(1417..1472,2781..2846,4783..4830,6240..6281,  
6552..6977,8611..8791)  
/gene="AMELX"  
/product="amelogenin precursor"  
1417..1472  
/gene="AMELX"  
/number=1  
2781..2846  
/gene="AMELX"  
/number=2  
join(2793..2846,4783..4830,6240..6281,6552..6977,  
8611..8616)  
/gene="AMELX"  
/codon\_start=1  
/product="amelogenin precursor"  
/protein\_id="AAK77213.1"  
/db\_xref="GI:15028583"  
/translation="MGTWILFACLLGAFAFMPLOPHRGHGYINFSYEVLTPLKWTOS  
MIRPP"

## exon

## exon

## CDS

join(1382..1437,3040..3051)  
/gene="AMELY"  
join(3052..3105,5192..5239,6461..>6505)  
/gene="AMELY"  
/product="amelogenin"  
/protein\_id="BAC6106.1"  
/db\_xref="GI:29126026"  
/translation="MGTWILFACLLGAFAFMPLOPHRGHGYINFSYEVLTPLKWTOS  
MIRPP"

sig\_peptide 2793..2840  
 /gene="AMELX"  
 exon 4783..4830  
 /gene="AMELX"  
 /number=3  
 exon 6107..6148  
 /gene="AMELX"  
 /number=4  
 exon 6240..6281  
 /gene="AMELX"  
 /number=5  
 exon 6552..6977  
 /gene="AMELX"  
 /number=6  
 exon 8611..8791  
 /gene="AMELX"  
 /number=7  
 polyA\_signal 8785..8791  
 /gene="AMELX"

Query Match 35.8%; Score 52.2; DB 9; Length 8810;  
 Best Local Similarity 76.6%; Pred. No. 0.00076;  
 Matches 82; Conservative 0; Mismatches 13; Indels 12; Gaps 1;  
 Oy 1 GATTTGTTGCGCCCTCGGAGCAGCCCTCAGTATGCTGTGATAAATTG--- 57  
 Db 2804 GATTTATTTGCTGCTCTCGGAGCAGCTTTGCGCATGCTGTGATAAACACCCCT 2863  
 Oy 58 -----CAATTCATTTTCACAGCTTGAAATMAAATCTGC 95  
 Db 2864 TGCATAGTCAGTGTCCATTTTCACAACTTGACATMAAATCTGC 2910

RESULT 11  
 AL357130/c 158142 bp DNA linear HTG 10-JUN-2001  
 LOCUS Homo sapiens chromosome X clone RP13-169E15, 4 unordered pieces.  
 AL357130.3 GI:9863814  
 ACCESSION AL357130.3  
 VERSION HTG; HTGS\_PHASE1; HTGS\_CANCELLED.  
 KEYWORDS Homo sapiens (human)  
 SOURCE Homo sapiens  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1  
 McIay, K.  
 Direct Submission  
 Submitted (09-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,  
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
 requests: clonerequest@sanger.ac.uk  
 On Aug 21, 2000 this sequence version replaced gi:9714076.  
 ----- Genome Center  
 Center: Sanger Centre  
 Center code: SC  
 Web site: http://www.sanger.ac.uk  
 Contact: humquery@sanger.ac.uk  
 ----- Project Information  
 Center project name: bbi69E15  
 ----- Summary Statistics  
 Assembly program: XGAP4; version 4.5  
 Sequencing vector: plasmid; L08752; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Consensus quality: 155448 bases at least Q40  
 Consensus quality: 156466 bases at least Q30  
 Consensus quality: 157038 bases at least Q20  
 Insert size: 157842; sum-of-contigs  
 Insert size: 160705; 33.4% error; agarose-fp  
 Quality coverage: 4.56x in Q20 bases; sum-of-contigs Quality  
 coverage: 4.56x in Q20 bases; agarose-fp  
 -----

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 4 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \* 1 68897: contig of 68897 bp in length  
 \* 68898 68997: gap of 100 bp  
 \* 68998 122842: contig of 53845 bp in length  
 \* 122843 122942: gap of 100 bp  
 \* 122943 129584: contig of 6642 bp in length  
 \* 129585 129684: gap of 100 bp  
 \* 129685 158142: contig of 28458 bp in length.  
 Location/Qualifiers  
 1..158142  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="X"  
 /clone="RP13-169E15"  
 /clone\_11b="RP13-13.1"  
 1..68897  
 /note="assembly fragment:01108  
 fragment chain:1  
 clone\_end:17  
 vector\_side:left"  
 68998..122842  
 /note="assembly fragment:00101  
 fragment chain:1"  
 122943..129584  
 /note="assembly fragment:00361  
 fragment chain:1"  
 129685..158142  
 /note="assembly fragment:01285  
 fragment chain:1  
 clone\_end:SP6  
 vector\_side:right"

misc\_feature  
 /note="assembly fragment:01108  
 fragment chain:1  
 clone\_end:17  
 vector\_side:left"  
 misc\_feature  
 /note="assembly fragment:00101  
 fragment chain:1"  
 122943..129584  
 /note="assembly fragment:00361  
 fragment chain:1"  
 129685..158142  
 /note="assembly fragment:01285  
 fragment chain:1  
 clone\_end:SP6  
 vector\_side:right"

Query Match 35.8%; Score 52.2; DB 2; Length 158142;  
 Best Local Similarity 76.6%; Pred. No. 0.00069;  
 Matches 82; Conservative 0; Mismatches 13; Indels 12; Gaps 1;  
 Oy 1 GATTTGTTGCGCCCTCGGAGCAGCCCTCAGTATGCTGTGATAAATTG--- 57  
 Db 44314 GATTTATTTGCTGCTCTCGGAGCAGCTTTGCGCATGCTGTGATAAACACCCCT 44255  
 Oy 58 -----CAATTCATTTTCACAGCTTGAAATMAAATCTGC 95  
 Db 44254 TGCATAGTCAGTGTCCATTTTCACAACTTGACATMAAATCTGC 44208

RESULT 12  
 AC002366 259202 bp DNA linear PRI 11-JUN-1998  
 LOCUS Human RP11-27C22 BAC CT-285115 (from CalTech/Research Genetics), PAC  
 DEFINITION RP11-27C22 (from Roswell Park Cancer Center), and Comid U5B5  
 (from Lawrence Livermore), complete sequence.  
 AC002366.1 GI:2739349  
 VERSION AC002366.1  
 KEYWORDS HTG.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 259202)  
 Muzny, D., Ansari-Lari, M.A., Timme, K.M., Yu, W., Dugan, S., Lu, J.,  
 Shen, Y., Rowland, K., Liu, W., Perez, L., Ding, Y., Haywood, M.,  
 Jain, A., Leal, B., Logan, O., Nguyen, V., Savage, L., Shen, H.,  
 Worley, K., Chen, E., Forcum, J., Atkinson, A.D., Chin, M.W.,  
 Gorrell, J.H., Brundage, E., Di, W., Chnault, C., Nelson, D. and  
 Gibbs, R.A.

TITLE	REFERENCE	AUTHORS	JOURNAL	REFERENCE	AUTHORS	JOURNAL	COMMENT
Direct Submission	Unpublished	2 (bases 1 to 259202)	Chiu,M.W.	Direct Submission	Submitted (23-JUN-1997)	Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA	3 (bases 1 to 259202)
Direct Submission	Submitted (02-JAN-1998)	Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA	4 (bases 1 to 259202)	Chiu,M.W.	Direct Submission	Submitted (11-JUN-1998)	Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Jan 2, 1998 this sequence version replaced by:2642176.	Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality does not meet this standard, it will be indicated in the annotation.	The repeat regions shown were identified using RepeatMasker by Adrian Smit.	Sequence similarities were identified using Powerblast by Jinghui Zhang.	Exon/Intron boundaries of identified genes were chosen if there were canonical splice junctions that maintained sequence continuity across the splice junctions.	Location/Qualifiers	1. 259202	/organism="Homo sapiens"
/mol_type="genomic DNA"	/db_xref="taxon:9606"	/chromosome="X"	/map="Xp22"	/clone="CT-285115, RPCC11-27C22, U5B5"	complement(1..147)	/note="overlap 1..147 in AC004467"	35..127
/rpt_family="MIR"	complement(1087..1502)	/rpt_family="t1"	complement(5292..5579)	/rpt_family="Alu"	5302..5553	/rpt_family="SVA"	complement(6009..6063)
/rpt_family="L1MC2"	complement(7092..7382)	/rpt_family="L1M45"	7423..7721	/rpt_family="Alu"	9771..10072	complement(10072..11146)	/rpt_family="THR"
complement(11156..11528)	/rpt_family="THE1D"	complement(12944..13151)	13862..14010	/rpt_family="MTR1f"	14063..14352	/rpt_family="MTR42a"	15847..16170
/rpt_family="Alu"	complement(18743..19117)	/rpt_family="MER7"	repeat_region	23507..23614	/rpt_family="L1"	23615..24425	/rpt_family="L1PA11"
complement(25475..25968)	/rpt_family="MTR1d"	27234..27683	/note="similar to EST with GenBank Accession Number H48827"	/db_xref="dbSTS:G289372"	complement(28186..28366)	/rpt_family="MER20"	28715..28995
/rpt_family="Alu"	complement(30217..30408)	31195..31364	/rpt_family="Alu"	complement(33939..34460)	34952..35384	/rpt_family="L1MA2"	36035..36126
/rpt_family="MTR1d"	36035..36126	/rpt_family="MIR2"	36850..37035	/rpt_family="MER20"	38042..38250	/db_xref="dbSTS:L47508"	38122..38510
/rpt_family="MSTa"	38197..40103	/rpt_family="MIR"	complement(42465..42539)	/rpt_family="MIR"	47135..47335	/rpt_family="MER3"	complement(48993..49166)
complement(49718..49918)	/rpt_family="L1PA11"	complement(51943..52505)	53226..53677	/rpt_family="L1MB3"	/rpt_family="MTR1d"	complement(54846..54966)	/rpt_family="MIR"
55040..55226	/rpt_family="MER20"	55691..55837	56051..56486	/rpt_family="L1MB3a"	58557..59030	/rpt_family="MTR1d"	complement(59270..59428)
/rpt_family="MERS"	complement(59847..60030)	/rpt_family="MER20"	complement(61290..61384)	61683..61811	/rpt_family="L1PB1"	61907..62113	/rpt_family="L1MB7"
complement(63586..63969)	64800..65073	complement(67157..67209)	67637..75845	/gene="HAMELX"	67637..68496	promoter	









**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 8, 2005, 11:25:10 ; Search time 193.675 Seconds  
(without alignments)  
3957.220 Million cell updates/sec

Title: US-10-754-437-11

Perfect score: 146  
Sequence: 1 gattctgttcgtcgtcc.....gtttaagacagtaacagatc 146

Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 413486 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

1: N\_Geneseq\_23Sep04:\*  
2: geneseqn1980s:\*  
3: geneseqn1990s:\*  
4: geneseqn2000s:\*  
5: geneseqn2001as:\*  
6: geneseqn2002as:\*  
7: geneseqn2002bs:\*  
8: geneseqn2003as:\*  
9: geneseqn2003bs:\*  
10: geneseqn2003cs:\*  
11: geneseqn2003ds:\*  
12: geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	52.2	35.8	259202	12	ADQ18492
2	40.2	27.5	722	6	AA14111
3	40.2	27.5	752	6	AA14110
4	38	26.0	270	12	ADM80835
5	38	26.0	549	12	ADM80826
6	38	26.0	623	12	ADM80827
7	36	24.7	825	4	AAH03775
8	36	24.7	1626	4	AAH17575
9	36	24.7	2176	4	AAH17101
10	36	24.7	3037	12	ADQ24117
11	34.6	23.7	556	10	ADBS9026
12	34.6	23.7	556	10	ADBS3782
13	34.4	23.6	149671	6	ABK84797
14	34.4	23.6	149671	9	ABD70361
15	34.4	23.6	149671	12	ADJ37140
16	34	23.3	65	6	ABN28389
17	34	23.3	476	3	AAZ50832
18	33.8	23.2	318	12	ADM80836
19	33.8	23.2	750	2	AAZ07020
20	33.8	23.2	793	2	AAZ07018
21	33.6	23.0	588	4	AAH10816

22	33.6	23.0	2497	4	AAH15986	Ah15986 Human CDN
23	32.6	22.3	5759	10	AD52555	Ad52555 Caenorhab
24	32.6	22.3	28564	10	ADP46508	Adp46508 Human gen
25	32.4	22.2	6081	6	ABL13429	Ab133429 Human imm
26	32.4	22.2	6081	6	ABN80193	Abn80193 Human che
27	32.4	22.2	15373	6	ABL32466	Ab132466 Human imm
28	32.2	22.1	802	2	AAZ07019	Aaz07019 Amelogeni
29	32.2	22.1	852	12	ADQ22977	Adq22977 Human sof
30	32	21.9	110000	12	ADN46845_17	Continuation (18 o
31	32	21.9	110000	12	ADN47591_02	Continuation (3 of
32	32	21.9	110000	12	ADN47591_03	Continuation (4 of
33	32	21.9	110000	12	ADN46123_17	Continuation (18 o
34	32	21.9	110000	12	ADN47209_02	Continuation (3 of
35	32	21.9	110000	12	ADN47209_03	Continuation (4 of
36	32	21.9	110000	12	ADN46464_17	Continuation (18 o
37	32	21.9	110000	12	ADN47960_02	Continuation (3 of
38	32	21.9	110000	12	ADN47960_03	Continuation (4 of
39	31.8	21.8	300	2	AAZ14959	Aaz14959 Human gen
40	31.8	21.8	4350	8	ABO83880	Abq83880 Human MDP
41	31.8	21.8	4410	8	ABO83860	Abq83860 Human MDP
42	31.8	21.8	6032	6	AA563310	Aa563310 Chemical1
43	31.8	21.8	16200	4	ABL13736	Ab13736 Drosophila
44	31.8	21.8	110000	2	AAZ42063_14	Continuation (15 o
45	31.4	21.5	440	8	ABX50288	Abx50288 Bovine ES

#### ALIGNMENTS

RESULT 1	ADQ18492	standard; DNA; 259202 BP.
ID	ADQ18492	
AC	ADQ18492;	
XX	26-AUG-2004 (first entry)	
DT	Human soft tissue sarcoma-upregulated DNA - SEQ ID 1311.	
XX	Human soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human; de.	
KM	Homo sapiens.	
OS	WO2004048938-A2.	
XX	10-JUN-2004.	
XX	26-NOV-2003; 2003WO-US038193.	
XX	26-NOV-2002; 2002US-0429739P.	
XX	(PROT-) PROTEIN DESIGN LABS INC.	
PA	Aziz N, Ginsburg WM, Zlotnick A;	
XX	WPI; 2004-441208/41.	
DR	Early detection of soft tissue sarcoma comprises determining expression of a gene in a first soft tissue sample and a normal soft tissue sample and comparing the gene expression, also useful in treating soft tissue sarcoma.	
PT	Example 2; SEQ ID NO 1311; 210bp; English.	
PS	The invention relates to a novel method for detecting soft tissue sarcoma which comprises obtaining a first soft tissue sample from an individual, and a normal soft tissue sample from the same or different individual, determining the expression of a gene in both samples and comparing the expression of the gene in both soft tissue samples, where a higher level of protein expression in the first soft tissue sample indicates the presence of soft tissue sarcoma. The method of the invention has cytostatic applications and may be useful for detecting soft tissue	



CC can be applied on different species of goats. This polynucleotide  
CC sequence represents a GAWL related X-chromosome DNA sequence of the  
CC invention

XX Sequence 752 BP, 184 A, 271 C, 152 G, 145 T, 0 U, 0 Other;

Query Match 27.5%; Score 40.2; DB 6; Length 752;  
Best Local Similarity 93.3%; Pred. No. 0.055; Indels 0; Gaps 0;  
Matches 42; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GATTTGTTGCTGCTCTGAGAGAGAGCTTCAGTATGCTGT 45  
DB 46 GATTTGTTGCTGCTCTGAGAGAGAGCTTCAGTATGCTGT 90

RESULT 4  
ADM80835  
ID ADM80835 standard; cDNA, 270 BP.

XX ADM80835;

DT 03-JUN-2004 (first entry)

XX Human CADECM-22 encoding cDNA SEQ ID NO:64.

DE human; cell adhesion and extracellular matrix protein; CADECM;  
KM neuroprotective; cytoskeletal; anorectic; immune disorder;  
KM neurological disorder; developmental disorder;  
KM connective tissue disorder; cell proliferative disorder; cancer; obesity;  
KM tangler disease; gene; ss.  
XX Homo sapiens.

OS  
XX Key Location/Qualifiers  
FH 36..212  
FT /\*tag= a  
FT /product= "CADECM-22"

XX WO2004015396-A2.

XX 19-FEB-2004.

XX 12-AUG-2003; 2003WO-US025418.

XX 13-AUG-2002; 2002US-0403781P.

XX 30-AUG-2002; 2002US-0407034P.

XX 13-SEP-2002; 2002US-0410566P.

XX 24-SEP-2002; 2002US-0413482P.

XX 25-SEP-2002; 2002US-0413890P.

XX 08-NOV-2002; 2002US-0424904P.

XX 13-NOV-2002; 2002US-0426222P.

XX (INCY-) INCYTE CORP.

XX Elloct VS, Khare R, Emerling BM, Kable AB, Tran UK, Jin P;

XX Becha SD, Margulis JP, Swarnaker A, Chawla NK, Ramkumar J;

XX Hafalia AJA, Lee SY, Jhang X, Jackson AA, Richardson TW, Blake JJ;

XX Wang JT, Chien D, Yang YG;

XX WPI; 2004-191795/18.

XX P-PSDB; ADM80793.

XX New cell adhesion and extracellular matrix proteins, useful in

XX diagnosing, treating and preventing immune, neurological, developmental,

XX connective tissue and cell proliferative disorders including cancer.

CC disorders including cancer, e.g. breast, prostate, ovarian, lung or colon  
CC cancer, obesity and tangler disease.

XX Sequence 270 BP, 76 A, 65 C, 67 G, 62 T, 0 U, 0 Other;

Query Match 26.0%; Score 38; DB 12; Length 270;  
Best Local Similarity 75.8%; Pred. No. 0.19; Indels 15; Gaps 0;  
Matches 47; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 1 GATTTGTTGCTGCTCTGAGAGAGAGCTTCAGTATGCTGTGAGTAAATTCGAA 60  
DB 47 GATTTATTTGCTGCTCTGAGAGAGAGCTTCAGTATGCTGTGAGTAAATTCGAA 106

QY 61 TT 62

DB 107 GT 108

RESULT 5  
ADM80826  
ID ADM80826 standard; cDNA, 549 BP.

XX ADM80826;

DT 03-JUN-2004 (first entry)

XX Human CADECM-13 encoding cDNA SEQ ID NO:55.

DE human; cell adhesion and extracellular matrix protein; CADECM;  
KM neuroprotective; cytoskeletal; anorectic; immune disorder;  
KM neurological disorder; developmental disorder;  
KM connective tissue disorder; cell proliferative disorder; cancer; obesity;  
KM tangler disease; gene; ss.  
XX Homo sapiens.

OS  
XX Key Location/Qualifiers  
FH 36..491  
FT /\*tag= a  
FT /product= "CADECM-13"

XX WO2004015396-A2.

XX 19-FEB-2004.

XX 12-AUG-2003; 2003WO-US025418.

XX 13-AUG-2002; 2002US-0403781P.

XX 30-AUG-2002; 2002US-0407034P.

XX 13-SEP-2002; 2002US-0410566P.

XX 24-SEP-2002; 2002US-0413482P.

XX 25-SEP-2002; 2002US-0413890P.

XX 08-NOV-2002; 2002US-0424904P.

XX 13-NOV-2002; 2002US-0426222P.

XX (INCY-) INCYTE CORP.

XX Elloct VS, Khare R, Emerling BM, Kable AB, Tran UK, Jin P;

XX Becha SD, Margulis JP, Swarnaker A, Chawla NK, Ramkumar J;

XX Hafalia AJA, Lee SY, Jhang X, Jackson AA, Richardson TW, Blake JJ;

XX Wang JT, Chien D, Yang YG;

XX WPI; 2004-191795/18.

XX P-PSDB; ADM80784.

XX New cell adhesion and extracellular matrix proteins, useful in

XX diagnosing, treating and preventing immune, neurological, developmental,

XX connective tissue and cell proliferative disorders including cancer.

XX Claim 5; SEQ ID NO 55; 272pp; English.

XX The present sequence encodes a human cell adhesion and extracellular

XX matrix protein designated CADECM. CADECM sequences has neuroprotective,



CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in the  
CC specification. The primer sets can be used in antisense therapy and in  
CC gene therapy. The primers are useful for synthesizing polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893  
CC represent human amino acid sequences; and AAH13629 to AAH13632 represent  
CC oligonucleotides, all of which are used in the exemplification of the  
CC present invention

SO Sequence 825 BP; 281 A; 123 C; 163 G; 255 T; 0 U; 3 Other;  
Query Match 24.7%; Score 36; DB 4; Length 825;  
Best Local Similarity 70.6%; Pred. No. 0.99;  
Matches 48; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

OY 4 TTTGTTGCTGCTCTCTGGAGCAGCCTTCAGTATGCTGTAGTAAATTCGAAATT 63  
DB 458 TTTCCTGACTGCTCATGAGACTAAATTCAGTTTCTGTATTACATTACACTT 399  
OY 64 CCAATTC 71  
DB 398 CTCCTTC 391

RESULT 8  
AAH17575/c  
ID AAH17575 standard; cDNA; 1626 BP.

AC AAH17575;  
XX 26-JUN-2001 (first entry)  
XX Human cDNA sequence SEQ ID NO:17065.  
DE Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
XX Homo sapiens.  
OS  
XX EP1074617-A2.  
XX 07-FEB-2001.  
XX 28-JUL-2000; 2000BP-00116126.  
XX 29-JUL-1999; 993P-00248036.  
XX 27-AUG-1999; 993P-00300253.  
XX 11-JAN-2000; 2000BP-00118776.  
XX 02-MAY-2000; 2000BP-00183767.  
XX 09-JUN-2000; 2000BP-00241899.  
XX (HELI-) HELIX RES INST.  
XX Ota T, Isegai T, Nishikawa T, Hayaehi K, Saito K, Yamamoto J,  
XX Iehi S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX WPI; 2001-318749/34.  
XX  
XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-  
XX length cDNAs defined in the specification, and for the detection and/or  
XX diagnosis of the abnormality of the proteins encoded by the full-length  
XX cDNAs.  
XX  
XX Claim 8; SEQ ID NO 17065; 2537pp + Sequence Listing; English.  
XX  
XX The present invention describes primer sets for synthesizing 5602 full-  
XX length cDNAs defined in the specification. Where a primer set comprises:  
XX (a) an oligo-dT primer and an oligonucleotide complementary to the  
XX complementary strand of a polynucleotide which comprises one of the 5602

CC nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in the  
CC specification. The primer sets can be used in antisense therapy and in  
CC gene therapy. The primers are useful for synthesizing polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893  
CC represent human amino acid sequences; and AAH13629 to AAH13632 represent  
CC oligonucleotides, all of which are used in the exemplification of the  
CC present invention

SO Sequence 1626 BP; 509 A; 297 C; 299 G; 521 T; 0 U; 0 Other;  
Query Match 24.7%; Score 36; DB 4; Length 1626;  
Best Local Similarity 70.6%; Pred. No. 1.2;  
Matches 48; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

OY 4 TTTGTTGCTGCTCTCTGGAGCAGCCTTCAGTATGCTGTAGTAAATTCGAAATT 63  
DB 913 TTTCCTGACTGCTCATGAGACTAAATTCAGTTTCTGTATTACATTACACTT 854  
OY 64 CCAATTC 71  
DB 853 CTCCTTC 846

RESULT 9  
AAH17101/c  
ID AAH17101 standard; cDNA; 2176 BP.

AC AAH17101;  
XX 26-JUN-2001 (first entry)  
XX Human cDNA sequence SEQ ID NO:16427.  
DE Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
XX Homo sapiens.  
OS  
XX EP1074617-A2.  
XX 07-FEB-2001.  
XX 28-JUL-2000; 2000BP-00116126.  
XX 29-JUL-1999; 993P-00248036.  
XX 27-AUG-1999; 993P-00300253.  
XX 11-JAN-2000; 2000BP-00118776.  
XX 02-MAY-2000; 2000BP-00183767.  
XX 09-JUN-2000; 2000BP-00241899.  
XX (HELI-) HELIX RES INST.  
XX Ota T, Isegai T, Nishikawa T, Hayaehi K, Saito K, Yamamoto J,  
XX Iehi S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX WPI; 2001-318749/34.  
XX  
XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-  
XX length cDNAs defined in the specification, and for the detection and/or  
XX diagnosis of the abnormality of the proteins encoded by the full-length  
XX cDNAs.  
XX  
XX Claim 8; SEQ ID NO 16427; 2537pp + Sequence Listing; English.  
XX

XX The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention.

XX Sequence 2176 BP; 683 A; 445 C; 394 G; 654 T; 0 U; 0 Other;

Query Match 24.7%; Score 36; DB 4; Length 2176;

Best Local Similarity 70.6%; Pred. No. 1.3; Mismatches 0; Indels 20; Gaps 0;

Matches 48; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 4 TTTGTTTGCTGCTGCTGAGAGAGAGCTTGTAGTATGCTGTGAGTAAATTTCAATTT 63  
DB 458 TTTCTTGAAGCTGCTGCTGAGAGAGTAAATTTCAATTTCTTCTTATTAACATTACACTTT 399

QY 64 CCAATTTTC 71  
DB 398 CTCCTTCC 391

#### RESULT 10

ADQ24117/C

ID ADQ24117 standard; DNA; 3037 BP.

XX ADQ24117;

DT 26-AUG-2004 (first entry)

XX Human soft tissue sarcoma-upregulated DNA - SEQ ID 6937.

XX soft tissue sarcoma; cytoskeletal; gene therapy; vaccine; screening; human; db.

OS Homo sapiens.

XX WO2004048938-A2.

XX 10-JUN-2004.

XX 26-NOV-2003; 2003WO-US038193.

XX 26-NOV-2002; 2002US-0429739P.

XX (PROT-) PROTEIN DESIGN LABS INC.

XX Aziz N, Ginsburg WM, Zlotnick A;

XX WPI, 2004-441208/41.

PT Early detection of soft tissue sarcoma comprises determining expression of a gene in a first soft tissue sample and a normal soft tissue sample and comparing the gene expression, also useful in treating soft tissue sarcoma.

PS Example 2; SEQ ID NO 6937; 210pp; English.

XX The invention relates to a novel method for detecting soft tissue sarcoma which comprises obtaining a first soft tissue sample from an individual and a normal soft tissue sample from the same or different individual, determining the expression of a gene in both samples and comparing the expression of the gene in both soft tissue samples, where a higher level of protein expression in the first soft tissue sample indicates the presence of soft tissue sarcoma. The method of the invention has cytostatic applications and may be useful for detecting soft tissue sarcoma, possibly via gene therapy or vaccine production. The nucleic acid sequences may be useful in diagnostic and screening applications. The current sequence is that of a human soft tissue sarcoma-upregulated DNA of the invention. The current sequence is not shown within the specification per se but was submitted in CD format by the inventor.

XX Sequence 3037 BP; 957 A; 596 C; 545 G; 919 T; 0 U; 20 Other;

Query Match 24.7%; Score 36; DB 12; Length 3037;

Best Local Similarity 70.6%; Pred. No. 1.4; Mismatches 0; Indels 20; Gaps 0;

Matches 48; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 4 TTTGTTTGCTGCTGCTGAGAGAGAGCTTGTAGTATGCTGTGAGTAAATTTCAATTT 63  
DB 913 TTTCTTGAAGCTGCTGCTGAGAGAGTAAATTTCAATTTCTTCTTATTAACATTACACTTT 854

QY 64 CCAATTTTC 71  
DB 853 CTCCTTCC 846

#### RESULT 11

ADB59026

ID ADB59026 standard; DNA; 556 BP.

XX ADB59026;

DT 04-DEC-2003 (first entry)

XX Toxicity-related gene, SEQ ID 4052.

XX Toxic; toxin; gene expression profile; hepatotoxicity; liver; drug screening; toxicity assay; db.

OS Unidentified.

XX WO2003064624-A2.

XX 07-AUG-2003.

XX 31-JAN-2003; 2003WO-US003194.

XX 31-JAN-2002; 2002US-00060087.

XX 15-MAR-2002; 2002US-0364045P.

XX 30-DEC-2002; 2002US-0436643P.

XX (GENE-) GENE LOGIC INC.

XX Mendrick D, Porter M, Johnson K, Higgs B, Castle A, Elashoff M;

XX WPI, 2003-689530/65.

PT Predicting a toxic effect of a compound, useful in identifying toxicity markers in liver tissues or cells for drug screening and toxicity assays, PT comprises preparing gene expression profile of tissue or cells exposed to the compound.

PS Claim 1; SEQ ID NO 4052; 1156pp; English.

XX The present invention relates to a method for predicting a toxic effect of a compound. The method comprises preparing a gene expression profile of a tissue or cell sample exposed to the compound, and comparing the









**This Page Blank (uspto)**



ZIP: 94304-1050  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch diskette  
COMPUTER: IBM compatible  
OPERATING SYSTEM: Microsoft Windows 95  
SOFTWARE: Wordperfect for windows 6.0,  
SOFTWARE: ASCII (DOS) TEXT format  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/018,595B  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: David J. Weitz  
REGISTRATION NUMBER: 38,362  
REFERENCE/DOCKET NUMBER: PELM-744  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 493-9300  
TELEFAX: (650) 493-6811  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 500 nucleotides  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-09-018-595B-3

Query Match 23.2%; Score 33.8; DB 2; Length 500;  
Best Local Similarity 84.4%; Pred. No. 0.26; 7; Indels 0; Gaps 0;  
Matches 38; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 GATTTGTTGCTGCTCTCTGGAGACAGCTTCAGTATGCTGT 45  
DB 80 GATTTATTGCTGCTCTCTGGAGACAGCTTTTGCATGCTCT 124

RESULT 3  
US-09-324-709A-3  
Sequence 3, Application US/09324709A  
Patent No. 6154707  
GENERAL INFORMATION:  
APPLICANT: Perkin-Elmer Corporation,  
APPLICANT: Applied Biosystems Division  
TITLE OF INVENTION: FLUORESCENCE GENOTYPING AT MULTIPLE ALLELIC  
SITES  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David J. Weitz,  
ADDRESSEE: Wilson Sonsini Goodrich & Rosati  
STREET: 650 Page Mill Road  
CITY: Palo Alto  
STATE: California  
COUNTRY: USA  
ZIP: 94304-1050  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch diskette  
COMPUTER: IBM compatible  
OPERATING SYSTEM: Microsoft Windows 95  
SOFTWARE: Wordperfect for windows 6.0,  
SOFTWARE: ASCII (DOS) TEXT format  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/324,709A  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: David J. Weitz  
REGISTRATION NUMBER: 38,362  
REFERENCE/DOCKET NUMBER: 16842-758

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 493-9300  
TELEFAX: (650) 493-6811  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 500 nucleotides  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-09-324-709A-3

Query Match 23.2%; Score 33.8; DB 3; Length 500;  
Best Local Similarity 84.4%; Pred. No. 0.26;  
Matches 38; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 GATTTGTTGCTGCTCTCTGGAGACAGCTTCAGTATGCTGT 45  
DB 80 GATTTATTGCTGCTCTCTGGAGACAGCTTTTGCATGCTCT 124

RESULT 4  
US-09-018-595B-1  
Sequence 1, Application US/09018595B  
Patent No. 5962233  
GENERAL INFORMATION:  
APPLICANT: Perkin-Elmer Corporation,  
APPLICANT: Applied Biosystems Division  
TITLE OF INVENTION: DETERMINATION OF GENOTYPE OF  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David J. Weitz,  
ADDRESSEE: Wilson Sonsini Goodrich & Rosati  
STREET: 650 Page Mill Road  
CITY: Palo Alto  
STATE: California  
COUNTRY: USA  
ZIP: 94304-1050  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch diskette  
COMPUTER: IBM compatible  
OPERATING SYSTEM: Microsoft Windows 95  
SOFTWARE: Wordperfect for windows 6.0,  
SOFTWARE: ASCII (DOS) TEXT format  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/018,595B  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: David J. Weitz  
REGISTRATION NUMBER: 38,362  
REFERENCE/DOCKET NUMBER: PELM-744  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 493-9300  
TELEFAX: (650) 493-6811  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 793 nucleotides  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-018-595B-1

Query Match 23.2%; Score 33.8; DB 2; Length 793;  
Best Local Similarity 84.4%; Pred. No. 0.31; 7; Indels 0; Gaps 0;  
Matches 38; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 GATTTGTTGCTGCTCTCTGGAGACAGCTTCAGTATGCTGT 45  
DB 80 GATTTATTGCTGCTCTCTGGAGACAGCTTTTGCATGCTCT 124

```
RESULT 5
US-09-324-709A-1
; Sequence 1, Application US/09324709A
; Patent No. 6154707
; GENERAL INFORMATION:
; APPLICANT: Perkin-Elmer Corporation,
; APPLICANT: Applied Biosystems Division
; TITLE OF INVENTION: FLUORESCENCE GENOTYPING AT MULTIPLE ALLELIC
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David J. Weitz,
; STREET: 650 Page Mill Road
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1050
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch diskette
; COMPUTER: IBM compatible
; OPERATING SYSTEM: Microsoft Windows 95
; SOFTWARE: Wordperfect for windows 6.0,
; SOFTWARE: ASCII (DOS) TEXT format
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/324,709A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: David J. Weitz
; REGISTRATION NUMBER: 38,362
; REFERENCE/DOCKET NUMBER: 16842-758
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 493-9300
; TELEFAX: (650) 493-6811
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 793 nucleotides
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-324-709A-1

Query Match      23.2%; Score 33.8; DB 3; Length 793;
Best Local Similarity 84.4%; Pred. No. 0.31;
Matches 38; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Cy      1 GATTGTGTCCTGCTGCTGAGAGAGAGCCTTCAGTATGCTCT 45
Db      80 GATTGTGTCCTGCTGCTGAGAGAGAGCCTTTGCCATGCTCT 124

RESULT 6
US-09-018-595B-2
; Sequence 2, Application US/09018595B
; Patent No. 5962233
; GENERAL INFORMATION:
; APPLICANT: Perkin-Elmer Corporation,
; APPLICANT: Applied Biosystems Division
; TITLE OF INVENTION: DETERMINATION OF GENOTYPE OF
; TITLE OF INVENTION: AMPLIFICATION PRODUCTS AT MULTIPLE ALLELIC SITES
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David J. Weitz,
; STREET: 650 Page Mill Road
; CITY: Palo Alto
; STATE: California
```

```
COUNTRY: USA
ZIP: 94304-1050
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: Microsoft Windows 95
SOFTWARE: Wordperfect for windows 6.0,
SOFTWARE: ASCII (DOS) TEXT format
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/018,595B
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: David J. Weitz
REGISTRATION NUMBER: 38,362
REFERENCE/DOCKET NUMBER: PELM-744
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 493-9300
TELEFAX: (650) 493-6811
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 802 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-018-595B-2

Query Match      22.1%; Score 32.2; DB 2; Length 802;
Best Local Similarity 82.2%; Pred. No. 0.95;
Matches 37; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Cy      1 GATTGTGTCCTGCTGCTGAGAGAGAGCCTTCAGTATGCTCT 45
Db      80 GATTGTGTCCTGCTGCTGAGAGAGAGCCTTTGCCATGCTCT 124

RESULT 7
US-09-324-709A-2
; Sequence 2, Application US/09324709A
; Patent No. 6154707
; GENERAL INFORMATION:
; APPLICANT: Perkin-Elmer Corporation,
; APPLICANT: Applied Biosystems Division
; TITLE OF INVENTION: FLUORESCENCE GENOTYPING AT MULTIPLE ALLELIC
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David J. Weitz,
; STREET: 650 Page Mill Road
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1050
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch diskette
; COMPUTER: IBM compatible
; OPERATING SYSTEM: Microsoft Windows 95
; SOFTWARE: Wordperfect for windows 6.0,
; SOFTWARE: ASCII (DOS) TEXT format
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/324,709A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: David J. Weitz
; REGISTRATION NUMBER: 38,362
```

REFERENCE/DOCKET NUMBER: 16842-758  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 493-9300  
TELEFAX: (650) 493-6811  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 802 nucleotides  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-324-709A-2

Query Match 22.1%; Score 32.2; DB 3; Length 802;  
Best Local Similarity 82.2%; Pred. No. 0.95;  
Matches 37; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 GATTTGTTGCTGCTCTCTGGAGCAGCTTCAGATGCTGT 45  
DB 80 GATTTGTTGCTGCTCTCTGGAGCAGCTTCAGATGCTGT 124

## RESULT 8

US-09-557-884-1  
Sequence 1, Application US/09557884  
Patent No. 6506581  
GENERAL INFORMATION:  
APPLICANT: Fleischmann et al.  
TITLE OF INVENTION: The Nucleotide sequence of the Haemophilus influenzae Rd Genome, Fragments thereof, and Uses Thereof  
NUMBER OF SEQUENCES: 1  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: MD  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3 1/2 inch diskette  
COMPUTER: Dell Pentium  
OPERATING SYSTEM: MS DOS v6.22  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/557,884  
FILING DATE: 25-Apr-2000  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/476,102  
FILING DATE: JUN-5-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Michelle S. Marks  
REGISTRATION NUMBER: 41,971  
REFERENCE/DOCKET NUMBER: PB186P3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 301-309-8504  
TELEFAX: 301-309-8439  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1830121 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-09-557-884-1

Query Match 21.8%; Score 31.8; DB 4; Length 1830121;  
Best Local Similarity 57.6%; Pred. No. 14;  
Matches 57; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 46 GAGTAAATTTCCAAATTTCCAAATTTCCAAAGCTTGAATTAATAATCTGCTCAGATTTT 105  
DB 1481806 GAGTAAATTTCCAAATTTTTTATCAAAAAGTATTAATAATAATCAAGCTTGACTTAT 1481865

QY 106 TATTTAGGTTAAATTTAAGGTTTAAAGCAGTACAGA 144  
DB 1481866 TCCTTTGCACTAGTTTAAATAGTACAAAATTTCCAAA 1481904

RESULT 9  
US-09-643-990A-1  
Sequence 1, Application US/09643990A  
Patent No. 6528289  
GENERAL INFORMATION:  
APPLICANT: Robert D. Adams  
Mark D. Adams  
Owen White  
Hamilton O. Smith  
J. Craig Venter

TITLE OF INVENTION: The Nucleotide sequence of the Haemophilus influenzae Rd Genome, Fragments thereof, and Uses Thereof

NUMBER OF SEQUENCES: 1  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville,  
STATE: MD  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3 1/2 inch diskette  
COMPUTER: Dell Pentium  
OPERATING SYSTEM: MS DOS v6.22  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/643,990A  
FILING DATE: 23-Aug-2000  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/487,429  
FILING DATE: 1995-06-07  
APPLICATION NUMBER: 08/426,787  
FILING DATE: 1995-04-21  
ATTORNEY/AGENT INFORMATION:  
NAME: Kenley K. Hoover  
REGISTRATION NUMBER: 40,302  
REFERENCE/DOCKET NUMBER: PB186P1C1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 301-610-5790  
TELEFAX: 310-309-8439  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1830121 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-09-643-990A-1

Query Match 21.8%; Score 31.8; DB 4; Length 1830121;  
Best Local Similarity 57.6%; Pred. No. 14;  
Matches 57; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 46 GAGTAAATTTCCAAATTTCCAAATTTCCAAAGCTTGAATTAATAATCTGCTCAGATTTT 105  
DB 1481806 GAGTAAATTTCCAAATTTTTTATCAAAAAGTATTAATAATAATCAAGCTTGACTTAT 1481865  
QY 106 TATTTAGGTTAAATTTAAGGTTTAAAGCAGTACAGA 144  
DB 1481866 TCCTTTGCACTAGTTTAAATAGTACAAAATTTCCAAA 1481904

RESULT 10  
US-10-323-960-1  
Sequence 1, Application US/10329960



```
/ Patent No. 6742927
/ GENERAL INFORMATION:
/ APPLICANT: Fleischmann et al.
/ TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus influenzae Rd Genome, Frag
/ Patent No. 6742927
/ TITLE OF INVENTION: Thereof, and Uses Thereof
/ FILE REFERENCE: P186P1
/ CURRENT APPLICATION NUMBER: US/10/329,960
/ CURRENT FILING DATE: 2003-01-02
/ PRIOR APPLICATION NUMBER: US 09/643,990
/ PRIOR FILING DATE: 2000-08-23
/ PRIOR APPLICATION NUMBER: US 08/487,429
/ PRIOR FILING DATE: 1995-06-07
/ PRIOR APPLICATION NUMBER: US 08/426,787
/ PRIOR FILING DATE: 1995-04-21
/ NUMBER OF SEQ ID NOS: 1
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 1
/ LENGTH: 1830121
/ TYPE: DNA
/ ORGANISM: Haemophilus influenzae
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (4747)..(4747)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (9921)..(9921)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (10150)..(10150)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (29298)..(29298)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (36543)..(36543)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (36551)..(36551)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (36636)..(36636)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (40808)..(40810)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (4416)..(4416)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (44905)..(44905)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (44975)..(44975)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (45593)..(45593)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (45732)..(45732)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (47036)..(47036)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (51309)..(51309)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (51334)..(51334)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (51602)..(51602)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (51786)..(51786)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (51805)..(51805)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (55369)..(55369)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (65309)..(65309)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (65313)..(65313)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (80024)..(80024)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (100091)..(100091)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (102696)..(102696)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (105121)..(105121)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (107248)..(107248)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (117136)..(117136)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (119750)..(119750)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (119924)..(119924)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (120038)..(120038)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (121344)..(121344)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
```

RESULT 13  
US-09-641-638-651  
; Sequence 651, Application US/09641638-651  
; Patent No. 6432648  
; GENERAL INFORMATION:  
; APPLICANT: Blumenfeld, Maria  
; APPLICANT: Bouguenelat, Lydie  
; APPLICANT: Chumakov, Ilya

```
APPLICANT: Cohen, Annick
TITLE OF INVENTION: BIALLERIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING
FILE REFERENCE: GENSET.051CPI
CURRENT APPLICATION NUMBER: US/09/641,638
PRIOR FILING DATE: 2000-08-16
PRIOR APPLICATION NUMBER: US 09/502,330
PRIOR FILING DATE: 2000-02-11
PRIOR APPLICATION NUMBER: US 60/133,200
PRIOR FILING DATE: 1999-05-07
PRIOR APPLICATION NUMBER: US 09/755,267
PRIOR FILING DATE: 1999-03-23
PRIOR APPLICATION NUMBER: US 60/119,917
PRIOR FILING DATE: 1998-02-12
NUMBER OF SEQ ID NOS: 1304
SOFTWARE: Patent.pm
SEQ ID NO 651
LENGTH: 20674
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1123..3123
OTHER INFORMATION: 5'regulatory region
NAME/KEY: exon
LOCATION: 3124..3297
OTHER INFORMATION: exon 1
NAME/KEY: exon
LOCATION: 3871..4072
OTHER INFORMATION: exon 2
NAME/KEY: exon
LOCATION: 5552..5633
OTHER INFORMATION: exon 3
NAME/KEY: exon
LOCATION: 5758..5880
OTHER INFORMATION: exon 4
NAME/KEY: exon
LOCATION: 5986..6099
OTHER INFORMATION: exon 5
NAME/KEY: exon
LOCATION: 6349..6509
OTHER INFORMATION: exon 6
NAME/KEY: exon
LOCATION: 7379..7522
OTHER INFORMATION: exon 7
NAME/KEY: exon
LOCATION: 8645..8854
OTHER INFORMATION: exon 8
NAME/KEY: exon
LOCATION: 12254..12340
OTHER INFORMATION: exon 9
NAME/KEY: exon
LOCATION: 12854..13023
OTHER INFORMATION: exon 10
NAME/KEY: exon
LOCATION: 13308..13429
OTHER INFORMATION: exon 11
NAME/KEY: exon
LOCATION: 16567..16667
OTHER INFORMATION: exon 12
NAME/KEY: exon
LOCATION: 16775..16945
OTHER INFORMATION: exon 13
NAME/KEY: exon
LOCATION: 17063..17554
OTHER INFORMATION: exon 14
NAME/KEY: misc_feature
LOCATION: 17555..20674
OTHER INFORMATION: 3'regulatory region
NAME/KEY: allele
LOCATION: 1128
OTHER INFORMATION: 10-508-191 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 1182
OTHER INFORMATION: 10-508-245 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 1559
OTHER INFORMATION: 10-509-284 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 1570
OTHER INFORMATION: 10-509-295 : deletion of C
NAME/KEY: allele
LOCATION: 1827
OTHER INFORMATION: 10-510-173 : variable motif ATTAA or TTTTTT
NAME/KEY: allele
LOCATION: 2048
OTHER INFORMATION: 10-511-62 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 2323
OTHER INFORMATION: 10-511-337 : insertion of T
NAME/KEY: allele
LOCATION: 2341
OTHER INFORMATION: 10-512-36 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 2623
OTHER INFORMATION: 10-512-318 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 2832
OTHER INFORMATION: 10-513-250 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 2844
OTHER INFORMATION: 10-513-262 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 2934
OTHER INFORMATION: 10-513-352 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 2947
OTHER INFORMATION: 10-513-365 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 3802
OTHER INFORMATION: 12-206-81 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 4062
OTHER INFORMATION: 10-343-231 : deletion of C
NAME/KEY: allele
LOCATION: 4088
OTHER INFORMATION: 12-206-366 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 4109
OTHER INFORMATION: 10-343-278 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 4170
OTHER INFORMATION: 10-343-339 : polymorphic base G or T
NAME/KEY: allele
LOCATION: 5903
OTHER INFORMATION: 10-346-23 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 6019
OTHER INFORMATION: 10-346-141 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 6141
OTHER INFORMATION: 10-346-263 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 6183
OTHER INFORMATION: 10-346-305 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 6338
OTHER INFORMATION: 10-347-74 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 6375
OTHER INFORMATION: 10-347-111 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 6429
OTHER INFORMATION: 10-347-165 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 6467
```

1. TITLE OF INVENTION: GENES INVOLVED IN ARACHIDONIC ACID METABOLISM

```

1 FILE REFERENCE: GEN-T114Xc2D1
2 CURRENT APPLICATION NUMBER: US/10/170,097
3 CURRENT FILING DATE: 2002-06-10
4 PRIOR APPLICATION NUMBER: US 09/641,638
5 PRIOR FILING DATE: 2000-08-16
6 PRIOR APPLICATION NUMBER: US 09/502,330
7 PRIOR FILING DATE: 2000-02-11
8 PRIOR APPLICATION NUMBER: US 60/133,200
9 PRIOR FILING DATE: 1999-05-07
10 PRIOR APPLICATION NUMBER: US 09/275,267
11 PRIOR FILING DATE: 1999-03-23
12 PRIOR APPLICATION NUMBER: US 60/119,917
13 PRIOR FILING DATE: 1999-02-12
14 NUMBER OF SEQ ID NOS: 1304
15 SOFTWARE: Patent.pm
16 SEQ ID NO 651
17 LENGTH: 20674
18 TYPE: DNA
19 ORGANISM: Homo sapiens
20 FEATURE:
21 NAME/KEY: misc feature
22 LOCATION: 1123..3123
23 OTHER INFORMATION: exon 1
24 FEATURE:
25 NAME/KEY: exon
26 LOCATION: 3871..4072
27 OTHER INFORMATION: exon 2
28 FEATURE:
29 NAME/KEY: exon
30 LOCATION: 5552..5633
31 OTHER INFORMATION: exon 3
32 FEATURE:
33 NAME/KEY: exon
34 LOCATION: 5758..5880
35 OTHER INFORMATION: exon 4
36 FEATURE:
37 NAME/KEY: exon
38 LOCATION: 5996..6099
39 OTHER INFORMATION: exon 5
40 FEATURE:
41 NAME/KEY: exon
42 LOCATION: 6349..6509
43 OTHER INFORMATION: exon 6
44 FEATURE:
45 NAME/KEY: exon
46 LOCATION: 7379..7522
47 OTHER INFORMATION: exon 7
48 FEATURE:
49 NAME/KEY: exon
50 LOCATION: 8645..8854
51 OTHER INFORMATION: exon 8
52 FEATURE:
53 NAME/KEY: exon
54 LOCATION: 12254..12340
55 OTHER INFORMATION: exon 9
56 FEATURE:
57 NAME/KEY: exon
58 LOCATION: 12854..13023
59 OTHER INFORMATION: exon 10
60 FEATURE:
61 NAME/KEY: exon
62 LOCATION: 13308..13429
63 OTHER INFORMATION: exon 11
64 FEATURE:
65 NAME/KEY: exon
66 LOCATION: 16567..16667
67 OTHER INFORMATION: exon 12
68 FEATURE:
69 NAME/KEY: exon

```

```
LOCATION: 16775..16945
OTHER INFORMATION: exon 13
FEATURE:
NAME/KEY: exon
LOCATION: 17063..17554
OTHER INFORMATION: exon 14
FEATURE:
NAME/KEY: misc feature
LOCATION: 17555..20674
OTHER INFORMATION: 3'regulatory region
FEATURE:
NAME/KEY: allele
LOCATION: 1128
OTHER INFORMATION: 10-508-191 : polymorphic base C or T
FEATURE:
NAME/KEY: allele
LOCATION: 1182
OTHER INFORMATION: 10-508-245 : polymorphic base C or T
FEATURE:
NAME/KEY: allele
LOCATION: 1559
OTHER INFORMATION: 10-509-284 : polymorphic base C or T
FEATURE:
NAME/KEY: allele
LOCATION: 1570
OTHER INFORMATION: 10-509-295 : deletion of C
FEATURE:
NAME/KEY: allele
LOCATION: 1827
OTHER INFORMATION: 10-510-173 : variable motif ATTTA or TTTTTT
FEATURE:
NAME/KEY: allele
LOCATION: 2048
OTHER INFORMATION: 10-511-62 : polymorphic base C or T
FEATURE:
NAME/KEY: allele
LOCATION: 2323
OTHER INFORMATION: 10-511-337 : insertion of T
FEATURE:
NAME/KEY: allele
LOCATION: 2341
OTHER INFORMATION: 10-512-36 : polymorphic base G or C
FEATURE:
NAME/KEY: allele
LOCATION: 2623
OTHER INFORMATION: 10-512-318 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 2832
OTHER INFORMATION: 10-513-250 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 2844
OTHER INFORMATION: 10-513-262 : polymorphic base C or T
FEATURE:
NAME/KEY: allele
LOCATION: 2934
OTHER INFORMATION: 10-513-352 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 2947
OTHER INFORMATION: 10-513-365 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 3802
OTHER INFORMATION: 12-206-81 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 4062
OTHER INFORMATION: 10-343-231 : deletion of C
FEATURE:
NAME/KEY: allele
LOCATION: 4088
```

```
OTHER INFORMATION: 12-206-366 : polymorphic base C or T
FEATURE:
NAME/KEY: allele
LOCATION: 4109
OTHER INFORMATION: 10-343-278 : polymorphic base C or T
FEATURE:
NAME/KEY: allele
LOCATION: 4170
OTHER INFORMATION: 10-343-339 : polymorphic base G or T
FEATURE:
NAME/KEY: allele
LOCATION: 5903
OTHER INFORMATION: 10-346-23 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 6019
OTHER INFORMATION: 10-346-141 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 6141
OTHER INFORMATION: 10-346-263 : polymorphic base G or C
FEATURE:
NAME/KEY: allele
LOCATION: 6183
OTHER INFORMATION: 10-346-305 : polymorphic base C or T
FEATURE:
NAME/KEY: allele
LOCATION: 6338
OTHER INFORMATION: 10-347-74 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 6375
OTHER INFORMATION: 10-347-111 : polymorphic base G or C
FEATURE:
NAME/KEY: allele
LOCATION: 6429
OTHER INFORMATION: 10-347-165 : polymorphic base C or T
FEATURE:
NAME/KEY: allele
LOCATION: 6467
OTHER INFORMATION: 10-347-203 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 6484
OTHER INFORMATION: 10-347-220 : polymorphic base A or G

Query Match          20.5%; Score 30; DB 4; Length 20674;
Best local similarity 59.3%; Pred. No. 13;
Matches 51; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 46 GAGTAAATTTCCATTTCACATTTCCACAGCTTGAAATTAATAATCTGCCTCAGATTTT 105
DB 11080 GAAATATTTTATATTTTAAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTT 11139
QY 106 TATTTAGGTTAAATTAAGCGTTA 131
DB 11140 AATTTAATATTAATTAATTAATTAATTTAA 11165

RESULT 15
US-08-916-421B-1
; Sequence 1, Application US/08916421B
; Patent No. 6503729
; GENERAL INFORMATION:
; APPLICANT: Buit et al.
; TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanoco-
; Patent No. 6503729
; TITLE OF INVENTION: Jannaeschli
; FILE REFERENCE: PB275
; CURRENT APPLICATION NUMBER: US/08/916,421B
; PRIOR APPLICATION NUMBER: 1997-08-22
; PRIOR FILING DATE: 1996-08-22
```

```
NUMBER OF SEQ ID NOS: 3
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1
LENGTH: 1664976
TYPE: DNA
ORGANISM: Mechanococcus jannaschii
FEATURE:
NAME/KEY: misc_feature
LOCATION: (28222)..(28222)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (28257)..(28258)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (84773)..(84773)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (84808)..(84808)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (84812)..(84812)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (98120)..(98120)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (98159)..(98159)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (98239)..(98239)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (98266)..(98266)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (98343)..(98343)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (10398)..(10398)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (148948)..(148948)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (163385)..(163385)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (191989)..(191989)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (191995)..(191995)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (231980)..(231980)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (234187)..(234187)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (234220)..(234220)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (234814)..(234814)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (309398)..(309398)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (309418)..(309418)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (312837)..(312837)
OTHER INFORMATION: n equals a, t, c, or g
```

```
NAME/KEY: misc_feature
LOCATION: (312993)..(312993)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (319226)..(319226)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (559167)..(559167)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (559241)..(559241)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (600992)..(600992)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (622708)..(622708)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (657081)..(657081)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (657203)..(657203)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (674435)..(674435)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (682442)..(682442)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (713652)..(713652)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (741684)..(741684)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (779455)..(779455)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (779676)..(779676)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (855539)..(855539)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (871619)..(871619)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1084830)..(1084830)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1096846)..(1096846)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1119881)..(1119881)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1130881)..(1130881)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1310988)..(1310988)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1313224)..(1313224)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1349473)..(1349473)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1349491)..(1349491)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
```

```
/ LOCATION: (1470091)..(1470091)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (1569020)..(1569020)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (1602912)..(1602912)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (1603734)..(1603734)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (1637998)..(1637998)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (1664854)..(1664855)
/ OTHER INFORMATION: n equals a, t, c, or g
US-08-916-421B-1
```

Query Match 20.54; Score 30; DB 4; Length 1664976;

Best Local Similarity 61.54; Pred. NO. 47; Matches 48; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

```
QY 68 TTTCACAGCTTGAAATAAAATCTGCTCAGATTTTATTAGGTTAAATTAAAGG 127
DB 1559577 TATGATGAGTTAGCATATATTTTAAAGACAGTTTATGAGATGTGATTGAAGG 1559636
QY 128 TTTAAGACAGTACAGAT 145
DB 1559637 GCTAAGATTTACTGAT 1559654
```

Search completed: February 8, 2005, 22:26:25  
Job time : 49.543 secs

**This Page Blank (uspto)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 8, 2005, 16:37:21 ; Search time 196.735 Seconds

(without alignments)  
4264.115 Million cell updates/sec

Title: US-10-754-437-11

Perfect score: 146  
Sequence: 1 gattctgttcgtcgtccctcc.....gtttaagacagacacagac 146

Scoring table: IDENTITY\_NNC  
Gapop 10.0, Gapext 1.0

Searched: 4300275 seqs, 2872944193 residues

Total number of hits satisfying chosen parameters: 8600550

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Published Applications NA:\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/2/pubpna/US09c\_PUBCOMB.seq:\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/2/pubpna/US09c\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09c\_PUBCOMB.seq:\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09c\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/2/pubpna/US09c\_PUBCOMB.seq:\*
- 12: /cgn2\_6/ptodata/2/pubpna/US09c\_PUBCOMB.seq:\*
- 13: /cgn2\_6/ptodata/2/pubpna/US10a\_PUBCOMB.seq:\*
- 14: /cgn2\_6/ptodata/2/pubpna/US10b\_PUBCOMB.seq:\*
- 15: /cgn2\_6/ptodata/2/pubpna/US10c\_PUBCOMB.seq:\*
- 16: /cgn2\_6/ptodata/2/pubpna/US10d\_PUBCOMB.seq:\*
- 17: /cgn2\_6/ptodata/2/pubpna/US10e\_PUBCOMB.seq:\*
- 18: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq:\*
- 19: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*
- 20: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*
- 21: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	52.8	36.2	600	US-10-027-632-287050	Sequence 287050,
C 2	52.8	36.2	600	US-10-027-632-287050	Sequence 287050,
C 3	52.2	35.8	600	US-10-027-632-287049	Sequence 287049,
C 4	52.2	35.8	600	US-10-027-632-287049	Sequence 287049,
C 5	52.2	35.8	600	US-10-027-632-287049	Sequence 287049,
C 6	52.2	35.8	600	US-10-027-632-287049	Sequence 287049,
C 7	35.8	24.7	18	US-10-723-860-1311	Sequence 6937, Ap
C 8	35	24.0	17	US-10-741-601-15294	Sequence 15294, A
C 9	34.4	23.6	15	US-10-236-031B-53	Sequence 5660, Ap
C 10	34.4	23.6	15	US-10-236-031B-53	Sequence 5660, Ap
C 11	33.8	23.2	65	US-09-908-975-1137	Sequence 1137, Ap
C 12	33.8	23.2	500	US-10-104-774-3	Sequence 3, Appli
				US-10-455-150-3	Sequence 3, Appli

13	33.8	23.2	793	13	US-10-104-774-1	Sequence 1, Appli
14	33.8	23.2	793	16	US-10-455-150-1	Sequence 1, Appli
15	33.8	23.2	793	17	US-10-755-889-605	Sequence 605, App
16	33.8	23.2	38299	17	US-10-741-601-5755	Sequence 5755, Ap
C 17	33.8	23.2	101782	17	US-10-741-601-5661	Sequence 5661, Ap
C 18	33.6	23.0	43871	17	US-10-741-601-5628	Sequence 5628, Ap
C 19	33.6	23.0	84886	17	US-10-741-601-5721	Sequence 5721, Ap
C 20	32.6	22.3	5759	16	US-10-347-470A-7	Sequence 7, Appli
C 21	32.4	22.2	6081	15	US-10-311-455-1402	Sequence 1402, Ap
C 22	32.4	22.2	15373	15	US-10-311-455-439	Sequence 439, App
C 23	32.2	22.1	802	13	US-10-104-774-2	Sequence 2, Appli
C 24	32.2	22.1	802	16	US-10-455-150-2	Sequence 2, Appli
C 25	32.2	22.1	852	18	US-10-723-860-5797	Sequence 5797, Ap
C 26	32	21.9	1773	16	US-10-424-599-135058	Sequence 135058,
C 27	32	21.9	289730	18	US-10-719-993-6780	Sequence 6780, Ap
C 28	31.8	21.8	3130	13	US-10-027-632-177165	Sequence 177165,
C 29	31.8	21.8	3130	15	US-10-027-632-177165	Sequence 177165,
C 30	31.8	21.8	4350	17	US-10-473-576-45	Sequence 45, Appli
C 31	31.8	21.8	4410	17	US-10-473-576-25	Sequence 25, Appli
C 32	31.8	21.8	6032	16	US-10-240-454-5	Sequence 5, Appli
C 33	31.8	21.8	1830121	14	US-10-328-960-1	Sequence 1, Appli
C 34	31.8	21.8	1830121	16	US-10-328-960-1	Sequence 1, Appli
C 35	31.8	21.8	1830121	18	US-10-158-865-1	Sequence 11205, A
C 36	31.6	21.6	1145	16	US-10-425-114-11205	Sequence 11205, A
C 37	31.6	21.6	1238	16	US-10-424-599-104040	Sequence 104040,
C 38	31.6	21.6	56155	17	US-10-741-601-5682	Sequence 5682, Ap
C 39	31.4	21.5	4450	9	US-09-983-965-217	Sequence 217, App
C 40	31.4	21.5	579	18	US-10-357-930-55270	Sequence 55270, A
C 41	31.2	21.4	650	13	US-10-027-632-226083	Sequence 226083,
C 42	31.2	21.4	650	13	US-10-027-632-226084	Sequence 226084,
C 43	31.2	21.4	650	15	US-10-027-632-226083	Sequence 226083,
C 44	31.2	21.4	650	15	US-10-027-632-226084	Sequence 226084,
C 45	31.2	21.4	672	16	US-10-424-599-113170	Sequence 113170,

#### ALIGNMENTS

RESULT 1  
US-10-027-632-287050/c  
; Sequence 287050, Application US/10027632  
; Publication No. US20020198371A1  
GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE REFERENCE: Polymorphisms in the Human Genome  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 287050  
; LENGTH: 600  
; TYPE: DNA  
; ORGANISM: Human  
; US-10-027-632-287050

Query Match 36.2%; Score 52.8; DB 13; Length 600;  
Best Local Similarity 75.9%; Pred. No. 5.9e-05;

```

OY      35.8%; Score 52.2; DB 13; Length 600;
Best Local Similarity 76.6%; Pred. No. 8.8e-05;
Matches 82; Conservative 0; Mismatches 13; Indels 12; Gaps 1;

OY      1 GATTTGTGGCTGCTCCTCTGGAGAGACCTTCAGTATGCTCTGTGATAAATTC--- 57
          |||||
Db      245 GATTTATTTGCTGCTCCTCTGGAGAGAGCTTTTGCCATGCCCTGTGATAAACCCTT 186
          |||||

OY      58 -----CAATTCATTTGACAAAGTTGGAATAAATATCTGC 95
          |||||
Db      185 TGCATTAAGTCAGTGTCCAAATTCACAACTTGGACATATAAATCTGC 139
          |||||

RESULT 4
US-10-027-632-287049/C
Sequence 287049, Application US/10027632
Publication No. US20030204075A9
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 287049
LENGTH: 600
TYPE: DNA
ORGANISM: Human
US-10-027-632-287049

OY      35.8%; Score 52.2; DB 13; Length 600;
Best Local Similarity 76.6%; Pred. No. 8.8e-05;
Matches 82; Conservative 0; Mismatches 13; Indels 12; Gaps 1;

OY      1 GATTTGTGGCTGCTCCTCTGGAGAGACCTTCAGTATGCTCTGTGATAAATTC--- 57
          |||||
Db      245 GATTTATTTGCTGCTCCTCTGGAGAGAGCTTTTGCCATGCCCTGTGATAAACCCTT 186
          |||||

OY      58 -----CAATTCATTTGACAAAGTTGGAATAAATATCTGC 95
          |||||
Db      185 TGCATTAAGTCAGTGTCCAAATTCACAACTTGGACATATAAATCTGC 139
          |||||

```



```

Query Match Similarity 23.6%; Score 34.4; DB 15; Length 149671;
Best Local Similarity 53.8%; Pred. No. 84;
Matches 71; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

OY 3 TTTGTTGGCTGCCTCCTGGAGCAGCCTTCAGATATGCTGTAGTAAATTTCCAAAT 62
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 33979 TTTAGTTGCTTGAACCCCTGTAACTCTCCACAGTACGTAGAGCTGTAAACACMACC 33920
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

OY 63 TCCATATTCACAGGCTGGAAATTAATAATCTGCTCAGATTTTATTAGGGTTAAATT 122
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 33919 TCCAGTTTCTTAAGCTAGTGTCTTAATAACTGAAGTCCTACTTTCTTAGTGATAGGG 33860
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

OY 123 AAGGTTTAAGA 134
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 33859 CAGGTTCCAGGA 33848
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 10
US-09-908-975-1137
; Sequence 1137, Application US/0908975
; Publication No. US20030165843A1
; GENERAL INFORMATION:
; APPLICANT: SHOSHAN, AVI
; APPLICANT: WASSERMAN, ALON
; APPLICANT: MINTZ, ELI
; APPLICANT: MINTZ, LIAT
; APPLICANT: FAISLER, SIMCHON
; TITLE OF INVENTION: OLIGONUCLEOTIDE LIBRARY FOR DETECTING RNA TRANSCRIPTS AND SPLICED
; TITLE OF INVENTION: THAT POPULATE A TRANSCRIPTOME
; FILE REFERENCE: 36688-0005
; CURRENT APPLICATION NUMBER: US/09/908,975
; CURRENT FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: US 60/287,724
; PRIOR FILING DATE: 2001-05-02
; PRIOR APPLICATION NUMBER: US 60/221,607
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 32337
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 1137
; LENGTH: 65
; TYPE: DNA
; ORGANISM: Rattus norvegicus
US-09-908-975-1137

Query Match Similarity 23.3%; Score 34; DB 10; Length 65;
Best Local Similarity 88.1%; Pred. No. 5.4;
Matches 37; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```

```
Query March      23.2%; Score 33.8; DB 13; Length 500;  
Best Local Similarity    86.4%; Pred. No. 14;  
Matches   38; Conservative    0; Mismatches     7; Indels     0; Gaps     0;
```

Db

          1   GATTTTGTCCTGCCTCCTGGAGACGCCCTTCAGTATCGCTCT  45  
            |||||                        |||  
      80   GATTTATTGCTGCCTCCTGGAGACGCTTTGGCATGCGCTCT  124  
            |||||                        |||

RESULT 12  
US-10-455-150-3  
Sequence 3, Application US/10455150  
Publication No. US20040053302A1  
GENERAL INFORMATION:  
APPLICANT: Perkin-Elmer Corporation,  
Applied Biosystems Division  
TITLE OF INVENTION: FLUORESCENCE GENOTYPING AT MULTIPLE ALLELIC  
SITES  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David J. Weitz,  
Wilson Sonsini Goodrich & Rosati  
STREET: 650 Page Mill Road  
CITY: Palo Alto  
STATE: California  
COUNTRY: USA  
ZIP: 94304-1050  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch diskette  
COMPUTER: IBM compatible  
OPERATING SYSTEM: Microsoft Windows 95  
SOFTWARE: Wordperfect for windows 6.0,  
ASCII (DOS) TEXT format  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/455,150  
FILING DATE: 04-Jun-2003  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/324,709A  
FILING DATE: 03-June-1999  
ATTORNEY/AGENT INFORMATION:  
NAME: David J. Weitz  
REGISTRATION NUMBER: 38,362  
REFERENCE/DOCKET NUMBER: 16842-758  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (493)-9300

```

; TELEFAX: (650) 493-6811
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 500 nucleotides
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-10-455-150-3

Query Match      23.2%; Score 33.8; DB 16; Length 500;
Best Local Similarity 84.4%; Pred. No. 14;
Matches 38; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 GATTTGTTGCTGCTGCTCTGAGAGAGAGCTTACGATGCTCT 45
DB 80 GATTTATTGCTGCTGCTCTGAGAGAGAGCTTTGCGATGCTCT 124

RESULT 13
US-10-104-774-1
; Sequence 1, Application US/10104774
; Publication No. US20020164630A1
; GENERAL INFORMATION:
; APPLICANT: Perkin-Elmer Corporation, Applied Biosystems Division
; TITLE OF INVENTION: DETERMINATION OF GENOTYPE OF AMPLIFICATION PRODUCTS AT MULTIPLE A
; FILE REFERENCE: 16842-782
; CURRENT APPLICATION NUMBER: US/10/104,774
; CURRENT FILING DATE: 2002-03-21
; PRIOR APPLICATION NUMBER: US 09/018,595
; PRIOR FILING DATE: 1998-02-04
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 793
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-104-774-1

Query Match      23.2%; Score 33.8; DB 13; Length 793;
Best Local Similarity 84.4%; Pred. No. 16;
Matches 38; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 GATTTGTTGCTGCTGCTCTGAGAGAGAGCTTACGATGCTCT 45
DB 80 GATTTATTGCTGCTGCTCTGAGAGAGAGCTTTGCGATGCTCT 124

RESULT 14
US-10-455-150-1
; Sequence 1, Application US/10455150
; Publication No. US20040053302A1
; GENERAL INFORMATION:
; APPLICANT: Perkin-Elmer Corporation,
; TITLE OF INVENTION: FLUORESCENCE GENOTYPING AT MULTIPLE ALLELIC
; SITES
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David J. Weitz,
;           Wilson Sonsini Goodrich & Rosati
;           STREET: 650 Page Mill Road
;           CITY: Palo Alto
;           STATE: California
;           COUNTRY: USA
;           ZIP: 94304-1050
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch diskette
; COMPUTER: IBM compatible
; OPERATING SYSTEM: Microsoft Windows 95
; SOFTWARE: Wordperfect for windows 6.0,
;           ASCII (DOS) TEXT format
```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/455,150
; FILING DATE: 04-Jun-2003
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/324,709A
; FILING DATE: 03-June-1999
; ATTORNEY/AGENT INFORMATION:
; NAME: David J. Weitz
; REGISTRATION NUMBER: 38,362
; REFERENCE/DOCKET NUMBER: 16842-758
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 493-9300
; TELEFAX: (650) 493-6811
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 793 nucleotides
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-10-455-150-1

Query Match      23.2%; Score 33.8; DB 16; Length 793;
Best Local Similarity 84.4%; Pred. No. 16;
Matches 38; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 GATTTGTTGCTGCTGCTCTGAGAGAGAGCTTACGATGCTCT 45
DB 80 GATTTATTGCTGCTGCTCTGAGAGAGAGCTTTGCGATGCTCT 124

RESULT 15
US-10-755-889-605
; Sequence 605, Application US/10755889
; Publication No. US20040171823A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ASSOCIATED WITH THE NF-KB
; FILE REFERENCE: D0284 NP
; CURRENT APPLICATION NUMBER: US/10/755,889
; CURRENT FILING DATE: 2004-01-13
; PRIOR APPLICATION NUMBER: U.S. 60/440,068
; PRIOR FILING DATE: 2003-01-14
; PRIOR APPLICATION NUMBER: U.S. 60/469,757
; PRIOR FILING DATE: 2003-05-12
; NUMBER OF SEQ ID NOS: 823
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 605
; LENGTH: 793
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-755-889-605

Query Match      23.2%; Score 33.8; DB 17; Length 793;
Best Local Similarity 84.4%; Pred. No. 16;
Matches 38; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 GATTTGTTGCTGCTGCTCTGAGAGAGAGCTTACGATGCTCT 45
DB 80 GATTTATTGCTGCTGCTCTGAGAGAGAGCTTTGCGATGCTCT 124

Search completed: February 9, 2005, 06:51:15
Job time : 203.735 secs
```

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using bw model

Run on: February 8, 2005, 12:58:25 ; Search time 1739 Seconds  
(without alignments)  
3059.349 Million cell updates/sec

Title: US-10-754-437-11

Perfect score: 146  
Sequence: 1 gattctgcttcgtcgtcc.....gttaagacagcacaagatc 146

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Database : Maximum Match 100%  
Listing first 45 summaries

EST:\*  
1: gb\_ests1:\*  
2: gb\_ests2:\*  
3: gb\_ests3:\*  
4: gb\_ests4:\*  
5: gb\_ests5:\*  
6: gb\_ests6:\*  
7: gb\_ests7:\*  
8: gb\_ests8:\*  
9: gb\_ests9:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	38.6	26.4	429	6	CB473334 sn68_E03.
2	37.4	25.6	633	6	CA682077 wlm24.pk0
3	36.2	24.8	1307	6	CG752765 P047-3-E0
4	36	24.7	331	5	BY496405 BY496405
5	36	24.7	441	1	AI798381 tr34c02.x
6	36	24.7	444	2	BE220076 hv66h03.x
7	36	24.7	448	1	AI338491 qg97h05.x
8	36	24.7	561	5	BA470683 DKF2P686N
9	36	24.7	586	2	BF001849 7994C09.x
10	36	24.7	588	2	BF448291 ncd17E04.
11	36	24.7	620	4	BG622884 602647718
12	36	24.7	707	4	CN360063 170005830
13	36	24.7	825	1	AU118731 AU118731
14	35.8	24.5	653	1	CE509949 t1gr-988
15	35.8	24.5	799	9	CL504547 SAIL_739
16	35.6	24.4	430	7	W33932 mb54d02.x1
17	35.6	24.4	430	7	W33932 mc43h12.x1
18	35.6	24.4	500	2	BE519442 MA000433
19	35.6	24.4	536	7	W36345 mb72c12.r1
20	35.6	24.4	585	9	AY419441 MWS muscu
21	35.6	24.4	645	6	CD773419 AGENCCOURT
22	35.6	24.4	707	6	CB056709 NISC_j119
23	35.6	24.4	812	3	AK029358 MWS muscu
24	35.6	24.4	843	6	CB588212 AGENCCOURT

25	35.6	24.4	859	6	CB588525	CB588525
26	35.6	24.4	865	6	CB589177	CB589177
27	35.6	24.4	866	6	CB587051	CB587051
28	35.6	24.4	889	6	CB589251	CB589251
29	35.6	24.4	891	6	CB574837	CB574837
30	35.6	24.4	913	6	CB590451	CB590451
31	35.6	24.4	919	6	CB587332	CB587332
32	35.6	24.4	928	6	CB590111	CB590111
33	35.6	24.4	945	2	BB614068	BB614068
34	35.6	24.4	990	7	W12906	W12906
35	35.6	24.4	1006	7	W29475	W29475
36	35.6	24.4	1020	7	W08102	W08102
37	35.4	24.2	496	5	B0741040	B0741040
38	35.4	24.2	600	8	AQ428965	AQ428965
39	35.2	24.1	689	8	B2017915	B2017915
40	35	24.0	633	4	B113076	B113076
41	35	24.0	872	8	BH167583	BH167583
42	35	24.0	893	8	AZ539059	AZ539059
43	35	24.0	904	8	AZ692144	AZ692144
44	34.8	23.8	478	8	AZ144107	AZ144107
45	34.8	23.8	965	9	CNS00302	AL064783 Drosophila

#### ALIGNMENTS

RESULT 1  
CB473334/c 429 bp mRNA linear EST 26-MAR-2003  
LOCUS sn68\_E03.f sn Sus scrofa cDNA 5', mRNA sequence.  
DEFINITION CB473334  
ACCESSION CB473334.1 GI:29279720  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Sus scrofa (pig)  
Sus scrofa  
Bukarya; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Theria; Cetartiodactyla; Suina; Suidae; Sus.

#### REFERENCE

#### TITLE

#### AUTHORS

#### ATLHORS

#### JOURNAL

#### COMMENT

#### CONTACT

#### PLUM

#### US

#### PO

#### Box

#### 848

#### Greenport

#### NY

#### 11944-848

#### USA

#### Tel:

#### 631 323 3133

#### Fax:

#### 631 323 3044

#### Email:

#### jnellan@pdc.ars.usda.gov

#### Single

#### pass

#### sequencing

#### Base

#### called

#### with

#### phred

#### v0.020425.c

#### and

#### trimmed

#### with

#### the

#### aid

#### of

#### the

#### trim

#### alt

#### option

#### Vector

#### identified

#### by

#### cross

#### match

#### v0.990329

#### and

#### Lucy

#### vi.17p

#### Seq

#### primer:

#### M13

#### Forward

#### Location/Qualifiers

#### 1..429

#### /organism="Sus scrofa"

#### /mol\_type="mRNA"

#### /db\_xref="taxon:9823"

#### /issue\_type="lymphoid"

#### /cell\_type="macrophage"

#### /lab\_host="DH10B"

#### /clone\_lib="sn"

#### /note="Vector: pSPORT1; Site.1: NotI; Site.2: SalI; Library made from pools of polyA selected RNA, isolated at different times post-infection (0 to 16 hrs) from African swine fever virus (isolate Pretoriuskop/96/4) infected swine macrophages. Macrophages were derived from peripheral blood mononuclear cells cultured for 48 hrs on plastic in the presence of 30% U937 supernatant."

#### ORIGIN

Query Match 26.4%; Score 38.6; DB 6; Length 429;  
Best local similarity 91.1%; Pred. No. 4.1;





Yang, L., Yuan, Z., Zavalan, M., Zhu, Y., Zimmer, A., Carrinci, P., Hayatsu, N., Hirozane-Kishikawa, T., Kono, H., Nakamura, M., Sakakuma, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishi, Y., Itoh, M., Kogawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shingawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S., Rogers, J., Birney, E. and Hayashizaki, Y.

Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

Nature 420, 563-573 (2002)

22354683

12466851

JOURNAL MEDLINE PUBLISHED COMMENT

Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gs.c.riken.jp, URL: http://genome.gsc.riken.jp/

Aizawa, K., Akimura, T., Arakawa, T., Carrinci, P., Fukuda, S., Hirozane, T., Imotani, K., Ishi, Y., Itoh, M., Kawai, J., Kono, H., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Waki, K., Watanabe, A., Muramatsu, M. and Hayashizaki, Y. Direct Submission

Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)

Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)

RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Tissues were provided by David A. Hume (Depts. of Biochemistry and Microbiology/Parasitology Institute for Molecular Bioscience University of Queensland Brisbane, Q 4072 Australia) whose assistance we gratefully acknowledge.

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

# FEATURES

source location/Qualifiers

1..331

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="1830045E16"

/issue\_type="bone marrow"

/cell\_type="macrophage"

/clone\_lib="RIKEN full-length enriched, bone marrow macrophage"

## ORIGIN

Query Match 24.7%; Score 36; DB 5; Length 331;  
Best Local Similarity 58.3%; Pred. No. 21;  
Matches 63; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

38 ATGCTGTGTGAGTAAATTTCCATTTTCACAGCTTGGAATAAATTCGCT 97  
|||||  
154 ATGCTGTGAGCTCAATGCGCATTAAGATTTTACCAACAAGTGGAAAAAGACTT 95  
|||||

QY 98 CAGATTTTATTTAGGCTTAATTTAAAGGTTTAAAGCAGTACAGAT 145  
|||||  
DB 94 CAAATTTTATTTCTATTTAGTTTAAATTTTAAATGAATATAT 47  
|||||

RESULT 5  
LOCUS A1798381 441 bp mRNA linear EST 06-JUL-1999  
DEFINITION tr34c02.x1 NCI\_CGAP\_Ov23 Homo sapiens cDNA clone IMAGE:2220194 3', mRNA sequence.

ACCESSION A1798381  
VERSION A1798381.1 GI:5363853  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 441)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapdb-remail.nih.gov  
Tissue Procurement: Christopher Mookajuk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: Life Technologies, Inc.  
DNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at: [www.bio.lnl.gov/bdrip/image/image.html](http://www.bio.lnl.gov/bdrip/image/image.html)  
Seq primer: -40UP from Glbco  
High quality sequence stop: 395.

# FEATURES

source location/Qualifiers

1..441

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:2220194"

/issue\_type="tumor, 5 pooled (see description)"

/lab\_host="DH10B"

/clone\_lib="NCI\_CGAP\_Ov23"

/note="Organ: ovary; Vector: pCMV-SPORT6; Site: 1; Salt; Site 2: Nci; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.35 kb. Tumor types include: mixed Mullerian tumor, papillary serous, clear cell, spindle cell. All are primary tumors, metastasis positive. Life Technologies catalog #: 11534-013"

## ORIGIN

Query Match 24.7%; Score 36; DB 1; Length 441;  
Best Local Similarity 70.6%; Pred. No. 22;  
Matches 48; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 4 TTGTGTCCTGCTCTCGGAGCAGCTTCAGTATCGTGTGAGTAAATTTCCATTT 63  
|||||  
DB 69 TTTCCTTGAAGCTGCTCATGAGCTTAATTCAGTTTCCGTATTAACATTTACACTT 128  
|||||

QY 64 CCAATTTTC 71  
|||||

DB 129 CTCCTTCC 136  
|||||

## ORIGIN

RESULT 6  
LOCUS BE220076 444 bp mRNA linear EST 03-JUL-2000  
DEFINITION hv66h03.x1 NCI\_CGAP\_Lu24 Homo sapiens cDNA clone IMAGE:3178421 3', mRNA sequence.

ACCESSION BE220076  
VERSION BE220076.1 GI:8907394  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 444)

**AUTHORS** NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
**TITLE** National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
**JOURNAL** Tumor Gene Index  
**COMMENT** Unpublished (1997)  
 Contact: Robert Strausberg, Ph.D.

Contact: Robert Strauberg, Ph.D.  
Email: cgaaps-r@mail.nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-buck, M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN, send email to: info@imgc.lnl.gov  
Seq primer: -40up from Gbbco  
High quality sequence, stop: 388.

FEATURES	Location/Qualifiers
source	1. .444

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3178421"
/tissue_type="carcinoid"
/lab_host="DH10b"
/clone_lib="NCI CGAP Lu24"
/note="Organ: Lung; Vector: pT773D-Pac (Pharmacia) with a
modified polylinker; Plasmid DNA from the normalized
library NCI CGAP Lu5 was prepared, and ss circles were
made in vitro. Following HAP purification, this DNA was
used as tracer in a subtractive hybridization reaction.
The driver was PCR-amplified cDNAs from a pool of 5,000
clones made from the same library (clonoids
1414920-1417991 and 1509094-1522439). Subtraction by Bencic
Soares and M. Fatima Bonaldo."

```

**ORIGIN**

Query Match	24.7%	Score 36;	DB 2;	Length 444;
Best Local Similarity	70.6%;	Pred. No. 22;		
Matches 48;	Conservative 0;	Mismatches 20;	Indels 0;	Gaps 0;

Dy            4 TTGTTTGCTGCCTCCTCGSAGACAGCTTCAGTATGCTGTGAATAATTTCCAATT 63  
             |||     |||     |||     |||     |||     |||     |||     |||     |||     |||  
Db            65 TTTCCTGACTGCCCATGAGAGCTAAATTCAGTTTTCTGTATTAACTATTACCTTT 122

Qy	64	CCAATTTC	71
Db	125	CTCCTTCC	132

RESULT 7	
A1338491	
LOCUS	
DEFINITION	A1338491 448 bp mRNA linear EST 13-FEB-1999
DESCRIPTION	gg97n05.x1 Soares_total_fetus Nb2HF8_9w Homo sapiens cDNA clone IMAGE:1939353 3', mRNA sequence.
CCDSSTON	12266401

ACCESSION	A1338491	
VERSION	A1338491.1	GI:4075418
KEYWORDS	EST.	
SOURCE	Homo sapiens (human)	
ORGANISM	Homo sapiens	

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 448)	NCI-CCGAP	<a href="http://www.ncbi.nlm.nih.gov/ncicgap">http://www.ncbi.nlm.nih.gov/ncicgap</a> .		
	National Cancer Institute,	Cancer Genome Anatomy Project (CGAP),		
	Tumor Gene Index	Unpublished (1997)		
	Contact: Robert Strausberg, Ph.D.			

Contact: Robert Strausberg, Ph.D.  
Email: c9abb8-r@mail.nih.gov  
This clone is available royalty-free through LNL, contact the  
IMAE Consortium ([info@mae.lnl.gov](mailto:info@mae.lnl.gov)) for further information.  
Insert Length: 1345 Std Error: 0.00  
Seq. primer: -400P from Gibco  
High quality sequence stop: 447.

**FEATURES**  
**Source**

```

1. 448
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mrna"
/db_xref="taxon:9606"
/clone="IMAGE:193953"
/dev_stage="8-9 weeks"
/lab_host="DH10B"
/clone_1ib="Soares total fetus Nb2H9.9w"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site 1: Not I, site 2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from pooled 8-9 week
total) fetus material with a Not I - 019G(dT) primer [5'].
TGTTCACATCTGAAGTGGACGCGCGCTTATATTTTATTTT 3'].
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo. "

```

**ORIGIN**

Query Match	24.7%	Score 36;	DB 1;	Length 448;
Best Local Similarity	70.6%	Pred. No. 22;		
Matches 48; Conservative	0;	Mismatches 20;	Indels 0;	Gaps 0;

0y 4TTTGTTCCTGCTCCTCGGACAGCCTTCAGATGCTGTGATGTTAAATTTCCAAATT 63  
Db 69 TTTTCTTACTGCTCAATGACCTAAATTCAGTTTCTGTATTTACATTTACACTTT 128

QY	64	CGAATTTC	71
Db	129	CTCCTTCC	136

LOCUS	DEFINITION	561 bp	mRNA	linear	EST 04-SEP-2003
BX470683/c	LOCUS	BX470683			
	DEFINITION	DKFZps868621110_r1 686 (synonym: hicc3) Homo sapiens CDNA clone DKFZps868621110 5', mRNA sequence.			

ACCESSION	BX470683	GI:3165009
VERSION	BX470683.1	
KEYWORDS	EST.	
SOURCE	Homo sapiens (human)	
ORGANISM	Homo sapiens	

REFERENCE  
AUTHORS  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
1 (bases 1 to 561)  
Koehnert, K., Beyer, A., Mewes, H.W., Weil, B., Amid, C., Osanger, A.,

TITLE EST (Koehler,K., Beyer,A., Mewes,H.W., weil,B., Amid,C., et al.,)  
JOURNAL Unpublished (2003)  
COMMENT Contact: MIPS

**FEATURES**

Ingolstaedter Landstr.1, D-85764 Neuberg, Germany  
This is the 5' sequence of the clone insert  
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer  
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;  
sequenced by BMFZ (Biomedical Research Center at the Heinrich-  
Heine-University, Duesseldorf/Germany) within the cDNA sequencing  
consortium of the German Genome Project.  
This clone (DKFZp68N2110) is available at the RZPD in Berlin.  
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059  
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.  
Location/Qualifiers

**FEATURES**  
**source**

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="DKFZP686N211.10"
/label="adult"
/lab_host="DH10B"
/clone_id="686 (synonym: hicc3)"
note="Vector: pSyn1Ex2; site_1: start; site_2: stop;"

```

# ORIGIN cdna-collection"

Query Match 24.7%; Score 36; DB 5; Length 561;  
Best Local Similarity 70.6%; Pred. No. 22;  
Matches 48; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 4 TTGTGTTGCTGCTCTGAGAGAGCCTTACAGTATGCTGTAGTAAATTTCCATTT 63  
DB 211 TTTCCTGACTGCTCATGAGAGCTAAATTCAGTTTCTGTATTAACTTTACACTTT 152

QY 64 CCAATTC 71  
DB 151 CTCCTTC 144

RESULT 9 BF001849 586 bp mRNA linear EST 06-OCT-2000  
LOCUS 7994609.x1 NCI\_CGAP\_Col6 Homo sapiens CDNA clone IMAGE:3314128 3',  
DEFINITION mRNA sequence.

ACCESSION BF001849 GI:10702124  
VERSION BF001849.1  
KEYWORDS EST.  
SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
1 (bases 1 to 586)

REFERENCE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)

JOURNAL Contact: Robert Strausberg, Ph.D.  
COMMENT Email: cgaaps-r@mail.nih.gov  
Tissue Procurement: Ian Kitzsch, M.D., Michael R. Emmert-Buck,  
M.D., Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D.  
CDNA Sequencing by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LINL, send email to:  
info@image.llnl.gov

Seq primer: -40UP from Gibco  
High quality sequence stop: 471.  
Location/Qualifiers

FEATURES  
source  
1..586  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3314128"  
/tissue\_type="colon tumor, RER+"  
/lab\_host="DH10B"  
/clone\_1b="NCI CGAP Col6"  
/note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a  
modified polylinker; Site 1: Not 1; Site 2: Eco RI;  
Plasmid DNA from the normalized library NCI\_CGAP\_Col6 was  
prepared, and 88 circles were made in vitro. Following HAP  
purification, this DNA was used as tracer in a subtractive  
hybridization reaction. The driver was PCR-amplified cDNAs  
from a pool of 5,000 clones made from the same library  
(clonoids 1057416-1061255, and 1144584-1145351).  
Subtraction by Bento Soares and M. Fatima Bonaldo. "

ORIGIN

Query Match 24.7%; Score 36; DB 2; Length 586;  
Best Local Similarity 70.6%; Pred. No. 22;  
Matches 48; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 4 TTGTGTTGCTGCTCTGAGAGAGCCTTACAGTATGCTGTAGTAAATTTCCATTT 63  
DB 65 TTTCCTGACTGCTCATGAGAGCTAAATTCAGTTTCTGTATTAACTTTACACTTT 124

QY 64 CCAATTC 71  
DB 125 CTCCTTC 132

RESULT 10 BF448291 588 bp mRNA linear EST 01-DEC-2000  
LOCUS nad17f04.x1 NCI\_CGAP\_Lu24 Homo sapiens CDNA clone IMAGE:3365887 3',  
DEFINITION mRNA sequence.

ACCESSION BF448291 GI:11513814  
VERSION BF448291.1  
KEYWORDS EST.  
SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
1 (bases 1 to 588)

REFERENCE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)

JOURNAL Contact: Robert Strausberg, Ph.D.  
COMMENT Email: cgaaps-r@mail.nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D.  
CDNA Sequencing by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LINL, send email to:  
info@image.llnl.gov

Seq primer: -40UP from Gibco  
High quality sequence stop: 465.  
Location/Qualifiers

FEATURES  
source  
1..588  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3365887"  
/tissue\_type="carcinoid"  
/lab\_host="DH10B"  
/clone\_1b="NCI CGAP Lu24"  
/note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a  
modified polylinker; Plasmid DNA from the normalized  
library NCI\_CGAP\_Lu5 was prepared, and 88 circles were  
made in vitro. Following HAP purification, this DNA was  
used as tracer in a subtractive hybridization reaction.  
The driver was PCR-amplified cDNAs from a pool of 5,000  
clones made from the same library (clonoids  
141920-1417991 and 1520904-1522439). Subtraction by Bento  
Soares and M. Fatima Bonaldo. "

ORIGIN

Query Match 24.7%; Score 36; DB 2; Length 588;  
Best Local Similarity 70.6%; Pred. No. 22;  
Matches 48; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 4 TTGTGTTGCTGCTCTGAGAGAGCCTTACAGTATGCTGTAGTAAATTTCCATTT 63  
DB 69 TTTCCTGACTGCTCATGAGAGCTAAATTCAGTTTCTGTATTAACTTTACACTTT 128

QY 64 CCAATTC 71  
DB 129 CTCCTTC 136

RESULT 11 BG622884/c 620 bp mRNA linear EST 18-APR-2001  
LOCUS 602647718F1 NIH\_MGC\_79 Homo sapiens CDNA clone IMAGE:4769213 5',  
DEFINITION mRNA sequence.

ACCESSION BG622884

VERSION EG622884.1 GI:13674255  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 620)  
 NIH-MGC <http://mgc.ncl.nih.gov/>.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
 Tissue Procurement: CLONTECH Laboratories, Inc.  
 cDNA Library Preparation: CLONTECH Laboratories, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: <http://image.llnl.gov>  
 Plate: L1CM1634 row: C column: 06  
 High quality sequence stop: 618.  
 Location/Qualifiers  
 1. 620  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4769213"  
 /lab\_host="DH10B (T1 phage-resistant)"  
 /clone\_1ib="NIH MGC 79"  
 /note="Organ: Placenta; Vector: pDNR-LIB (Clontech); site\_1: sfil (ggcgctggcc); site\_2: sfil (ggcgctggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGGCCATTATGCGC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGCGCGCCGACATG-dt(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.3 kb (range 0.5-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH\_MGC Library."

ORIGIN  
 Query Match 24.7%; Score 36; DB 4; Length 620;  
 Best Local Similarity 70.6%; Pred. No. 23;  
 Matches 48; Conservative 0; Mismatches 20; Indels 0; Gaps 0;  
 QY 4 TTGTGTTGCTGCTCTCGGAGCAGCCTTCAGTATGCTGTGAGTAAATTTTCAATT 63  
 DB 513 TTTCCTGACCTGCTCATGAGAGCTAAATTCAGTTTCTGTATTAACTTTACACTT 454  
 QY 64 CCAATTTC 71  
 DB 453 CTCCTTCC 446

RESULT 12  
 CN360063/c 707 bp mRNA linear EST 16-MAY-2004  
 LOCUS 17000583090269 GRN\_PPREHP Homo sapiens cDNA 5', mRNA sequence.  
 ACCESSION CN360063  
 VERSION CN360063.1 GI:47359997  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 707)  
 Brandenberger, R., Wei, H., Zhang, S., Lei, S., Murage, J., Fisk, G.J., Li, Y., Xu, C., Fang, R., Guegler, K., Rao, M.S., Mandalam, R., Lebkowski, J. and Stanton, L.W.  
 Transcriptional characterization elucidates signaling networks that control human ES cell growth and differentiation  
 Nat. Biotechnol. 22 (6), 707-716 (2004)

COMMENT Contact: Brandenberger R  
 Regenerative Medicine  
 Genon Corporation  
 230 Constitution Drive, Menlo Park, CA 94025, USA  
 Tel: 650 473 8658  
 Fax: 650 473 7760  
 Email: [rbrandenberger@genon.com](mailto:rbrandenberger@genon.com)  
 Insert length: 707 Std Error: 0.00.  
 Location/Qualifiers  
 1. 707  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /tissue\_type="embryonic stem cells, DMSO-treated H9 cell line"  
 /clone\_1ib="GRN\_PPREHP"  
 /note="oligo dt primed, full-length enriched cDNA library from DMSO-treated hES cell line H9 (p22) maintained in feeder-free conditions"

ORIGIN  
 Query Match 24.7%; Score 36; DB 7; Length 707;  
 Best Local Similarity 70.6%; Pred. No. 23;  
 Matches 48; Conservative 0; Mismatches 20; Indels 0; Gaps 0;  
 QY 4 TTGTGTTGCTGCTCTCGGAGCAGCCTTCAGTATGCTGTGAGTAAATTTTCAATT 63  
 DB 596 TTTCCTGACCTGCTCATGAGAGCTAAATTCAGTTTCTGTATTAACTTTACACTT 537  
 QY 64 CCAATTTC 71  
 DB 536 CTCCTTCC 529

RESULT 13  
 AU118731/c 825 bp mRNA linear EST 01-AUG-2002  
 LOCUS AU118731 HEMBA1 Homo sapiens cDNA clone HEMBA1004241 5', mRNA  
 DEFINITION AU118731  
 ACCESSION AU118731  
 VERSION AU118731.1 GI:10933870  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 825)  
 Oca, T., Nishikawa, T., Suzuki, Y., Ishii, S., Saito, K., Kawai, Y., Yamamoto, J., Wakamatsu, A., Nakamura, Y., Nagai, T., Sugano, S. and Isogai, T.  
 HRI human cDNA project  
 Unpublished (2000)  
 Contact: Takao Isogai  
 Genomics Laboratory  
 Helix Research Institute  
 1532-3 Yana, Kisarazu, Chiba 229-0812, Japan  
 Tel: 81-438-52-3975  
 Fax: 81-438-52-3986  
 Email: [genomics@hri.co.jp](mailto:genomics@hri.co.jp)  
 HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix Research Institute; cDNA library construction: Department of Virology, Institute of Medical Science, University of Tokyo, and Helix Research Institute.  
 Location/Qualifiers  
 1. 825  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="HEMBA1004241"  
 /tissue\_type="whole embryo, mainly head"  
 /dev\_stage="embryo, 10 weeks"  
 /clone\_1ib="HEMBA1"  
 /note="Vector: PME18SFL3"

## ORIGIN

Query Match 24.7%; Score 36; DB 1; Length 825;  
 Best Local Similarity 70.6%; Pred. No. 24;  
 Matches 48; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 4 TTTGTTGCTGCTGCTGAGAGAGAGCTTCAGTATCGGTGAGTAATTTCCATT 63  
 DB 458 TTTCTTGACTGCTCATGAGAGCTAAATTCAGTTTCTGTTATTAACATTTACCTT 399

QY 64 CCAATTTG 71  
 DB 398 CTCCTTCC 391

RESULT 14  
 CES09949/c 653 bp DNA linear GSS 28-SEP-2003  
 LOCUS tigr-gss-dog-17000310716522 Dog Library Canis familiaris genomic,  
 DEFINITION genomic survey sequence.

ACCESSION CES09949  
 VERSION CES09949.1 GI:36826730  
 KEYWORDS GSS.  
 SOURCE Canis familiaris (dog)  
 ORGANISM Canis familiaris

REFERENCE 1 (bases 1 to 653)  
 Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K.,  
 Ruesch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and  
 Venter,J.C.

TITLE The dog genome: survey sequencing and comparative analysis  
 JOURNAL Science 301 (5641), 1898-1903 (2003)  
 MEDLINE 22875432  
 PUBMED 14512627

COMMENT Contact: Kirkness EF  
 The Institute for Genomic Research  
 Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,  
 Rockville, MD 20850, USA  
 Tel: 301-838-0200  
 Fax: 301-838-0208  
 Email: ekirkness@tigr.org  
 Class: shotgun.

## FEATURES

source location/Qualifiers  
 1..653  
 /organism="Canis familiaris"  
 /mol\_type="genomic DNA"  
 /strain="Standard Poodle"  
 /db\_xref="taxon:9615"  
 /clone\_lib="Dog Library"  
 /note="Site 1: BcXI; Libraries were prepared from  
 peripheral blood"

## ORIGIN

Query Match 24.5%; Score 35.8; DB 9; Length 653;  
 Best Local Similarity 59.2%; Pred. No. 26;  
 Matches 61; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 30 CCTTCAGTATGCTGTGAGTAATTTCCATTTCACAGCTTGAATTAATAA 89  
 DB 653 CTTTCATTATGATATGATATGATATTAACAGTTTCAATATATATAGAGAAATATAT 594

QY 90 ATCTGCCTCAGATTTTATTTAGGCTTAATTTAAGGCTTTAA 132  
 DB 593 GCATTACACAGATTGCTTTTCTTTTCTTTTAAAGATTTTA 551

RESULT 15  
 CL504547 799 bp DNA linear GSS 01-APR-2004  
 LOCUS SAIL\_739\_G07.v1 SAIL Collection Arabidopsis thaliana genomic clone  
 DEFINITION SAIL\_739\_G07.v1, genomic survey sequence.  
 ACCESSION CL504547

## VERSION

CL504547.1 GI:46001867  
 GSS.  
 Arabidopsis thaliana (thale cress)

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

Sessions,A., Burke,E., Presting,G., Aux,G., McElver,J., Patton,D.,  
 Dietrich,B., Ho,P., Bacwaden,J., Ko,C., Clarke,D., Cotton,D.,  
 Bullis,D., Snell,J., Miguel,T., Hutchison,D., Kimmery,B.,  
 Mitzel,T., Katagiri,F., Glazebrook,J., Law,M. and Goff,S.A.  
 A high-throughput Arabidopsis reverse genetics system  
 Plant Cell 14 (12), 2985-2994 (2002)

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## COMMENT

Contact: Sessions A  
 Applied Trait Genetics  
 Syngenta Biotechnology Inc.  
 3054 Cornwallis Rd., Research Triangle Park, NC 27709, USA  
 Email: allen.sessions@syngenta.com  
 ABRC Stock Number CS833063; T-DNA left border flanking sequences of  
 Syngenta Arabidopsis insertion library (SAIL) lines are available  
 through the Arabidopsis Biological Resource Center (ABRC).  
 Sequences represent a pool of amplified genomic regions and not  
 single contiguous sequences.  
 Class: T-DNA tagged.

## FEATURES

## source

location/Qualifiers  
 1..799  
 /organism="Arabidopsis thaliana"  
 /mol\_type="genomic DNA"  
 /ecotype="Columbia"  
 /db\_xref="taxon:3702"  
 /clone\_lib="SAIL\_739\_G07.v1"  
 /note="T-DNA left border sequences were isolated using a  
 modified TAIL-PCR strategy"

## ORIGIN

Query Match 24.5%; Score 35.8; DB 9; Length 799;  
 Best Local Similarity 59.2%; Pred. No. 27;  
 Matches 61; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 30 CCTTCAGTATGCTGTGAGTAATTTCCATTTCACAGCTTGAATTAATAA 89  
 DB 496 CATTATATATCTTTTGAAAAAATTTGATTTGATTTGCAACCTTTTAAAAAA 555

QY 90 ATCTGCCTCAGATTTTATTTAGGCTTAATTTAAGGCTTTAA 132  
 DB 556 AAAAAAAGATATTTTCTGCGGCCCTTAAGGGCGCAA 598

Search completed: February 8, 2005, 22:18:20  
 Job time : 1742 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 8, 2005, 11:28:00 ; Search time 2743.23 Seconds  
(without alignments)  
8481.439 Million cell updates/sec

Title: US-10-754-437-22

Perfect score: 492  
Sequence: 1 atggggagactgattgtt.....aactctctccgcctcca 492

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Genbank:.\*  
1: gb\_ba:.\*  
2: gb\_ncg:.\*  
3: gb\_in:.\*  
4: gb\_cm:.\*  
5: gb\_ov:.\*  
6: gb\_pat:.\*  
7: gb\_ph:.\*  
8: gb\_dl:.\*  
9: gb\_pr:.\*  
10: gb\_ro:.\*  
11: gb\_ste:.\*  
12: gb\_sy:.\*  
13: gb\_un:.\*  
14: gb\_vt:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	339	66.9	5712	4	AB091793 Equus cab
2	303.2	61.6	6451	4	AB091789 Bos tauru
3	302.8	61.5	7425	4	AB091791 Sus scrof
4	274.6	55.8	8810	9	AY040206 Homo sapi
5	274.6	55.8	158142	2	AL357130 Homo sapi
6	274.6	55.8	259202	9	AC002366 Human Xp2
7	273	55.5	6465	9	AB091781 Pan trogl
8	261	53.0	5684	9	AB091785 Lemur cal
9	256.6	52.2	6442	9	AB091783 Saimiri s
10	248.8	50.6	5591	4	AB091794 Equus cab
11	245.2	49.8	5562	9	AB091787 Ocolemur
12	219.6	44.6	1935	9	D83730 Homo sapien
13	166.2	33.8	6264	4	AB091790 Bos tauru
14	160.6	32.6	9384	10	AF294397 Mus muscu
15	160.6	32.6	95826	10	AL805974 Mus muscu
16	150	30.5	695	11	BV089295 RPPAMSE00
17	150	30.5	695	11	BV097603 RPPAMSE00
18	142.2	28.9	212886	2	AC093946 Rattus norv
19	142.2	28.9	234471	2	AC121424 Rattus norv

20	127.6	25.9	8004	9	AB091786 Lemur cat
21	116.2	23.6	7454	9	AB091784 Saimiri s
22	96.8	19.7	463	10	D8306382
23	95.4	19.4	177654	2	AP000918
24	95.4	19.4	200214	9	AC013412 Homo sapi
25	94.8	19.3	7163	9	AB091782 Pan trogl
26	94.8	19.3	38765	9	BS000568 Pan trogl
27	94.8	19.3	177726	9	AC145770 Pan trogl
28	94.8	19.3	190089	9	BS000576 Pan trogl
29	93	18.9	6931	4	AB091792 Sus scrof
30	82.8	16.8	2501	9	D83729
31	73	14.8	363	11	BV089294 RPPAMSE00
32	73	14.8	363	11	BV097602 RPPAMSE00
33	52.8	10.7	226	10	S74899
34	52.8	10.7	727	10	MUSMAMEB
35	52.8	10.7	765	4	AB032194
36	52.8	10.7	789	4	AB032193
37	52.8	10.7	798	10	BC059090
38	52.8	10.7	799	10	MUSMAMEA
39	51.6	10.5	166426	9	AC025505
40	51.6	10.5	187161	9	AC090824
41	51.2	10.4	399	10	RR07054
42	51.2	10.4	457	10	RNU60564
43	51.2	10.4	476	6	AR452534
44	51.2	10.4	753	10	RNU60562
45	51.2	10.4	780	10	U01245

#### ALIGNMENTS

RESULT 1  
AB091793  
LOCUS  
DEFINITION  
AB091793  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
AB091793  
Equus caballus AMELX gene for amelogenin, partial cds.  
AB091793.1 GI:29126038

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
12672962  
2 (bases 1 to 5712)  
Iwase,M., Saita,Y. and Takahata,N.  
Submitted (19-SEP-2002) Mineyo Iwase, Graduate University for  
Advanced Studies (Soken-dai), Department of Biosystems Science;  
Shonan kokusai-mura, Hayama, Kanagawa 240-0193, Japan  
(E-mail:iwase@min@koryuwo1.soken.ac.jp, Tel:81-468-58-1571,  
Fax:81-468-58-1544)  
Location/Qualifiers  
1..5712  
/organism="Equus caballus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9796"  
/sex="male"  
174..5712  
/gene="AMELX"  
join(174..233,1578..1589)  
/gene="AMELX"  
join(1590..1643,3566..3613,4993..5037,5305..55712)  
/gene="AMELX"  
/codon\_start=1  
/product="amelogenin"  
/protein\_id="BAC66112.1"  
/db\_xref="GI:29126039"

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
12672962  
2 (bases 1 to 5712)  
Iwase,M., Saita,Y. and Takahata,N.  
Submitted (19-SEP-2002) Mineyo Iwase, Graduate University for  
Advanced Studies (Soken-dai), Department of Biosystems Science;  
Shonan kokusai-mura, Hayama, Kanagawa 240-0193, Japan  
(E-mail:iwase@min@koryuwo1.soken.ac.jp, Tel:81-468-58-1571,  
Fax:81-468-58-1544)  
Location/Qualifiers  
1..5712  
/organism="Equus caballus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9796"  
/sex="male"  
174..5712  
/gene="AMELX"  
join(174..233,1578..1589)  
/gene="AMELX"  
join(1590..1643,3566..3613,4993..5037,5305..55712)  
/gene="AMELX"  
/codon\_start=1  
/product="amelogenin"  
/protein\_id="BAC66112.1"  
/db\_xref="GI:29126039"

#### FEATURES

source  
1..5712  
/organism="Equus caballus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9796"  
/sex="male"  
174..5712  
/gene="AMELX"  
join(174..233,1578..1589)  
/gene="AMELX"  
join(1590..1643,3566..3613,4993..5037,5305..55712)  
/gene="AMELX"  
/codon\_start=1  
/product="amelogenin"  
/protein\_id="BAC66112.1"  
/db\_xref="GI:29126039"





SOURCE	Sus scrofa (pig)
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE AUTHORS TITLE	Iwase, M., Satta, Y., Hitai, Y., Hitai, H., Imai, H. and Takahara, N. From the Cover: The amelogenin loci span an ancient pseudautosomal boundary in diverse mammalian species <i>Proc. Natl. Acad. Sci. U.S.A.</i> 100 (9), 5258-5263 (2003)
JOURNAL MEDLINE PUBMED	12672962 2 (bases 1 to 7425) Iwase, M., Satta, Y. and Takahara, N. Direct Submission Submitted (19-SEP-2002) Mineyo Iwase, Graduate University for Advanced Studies (Sokenaki), Department of Biosystems Science, Shonan Kikusaimura, Hayama, Kanagawa 240-0193, Japan (E-mail:iwasemin@koryu.w01.boken.ac.jp, Tel:81-468-58-1571, Fax:81-468-58-1544)
REFERENCE AUTHORS TITLE	Iwase, M., Satta, Y., Hitai, Y., Hitai, H., Imai, H. and Takahara, N. From the Cover: The amelogenin loci span an ancient pseudautosomal boundary in diverse mammalian species <i>Proc. Natl. Acad. Sci. U.S.A.</i> 100 (9), 5258-5263 (2003)
JOURNAL MEDLINE PUBMED	12672962 2 (bases 1 to 7425) Iwase, M., Satta, Y. and Takahara, N. Direct Submission Submitted (19-SEP-2002) Mineyo Iwase, Graduate University for Advanced Studies (Sokenaki), Department of Biosystems Science, Shonan Kikusaimura, Hayama, Kanagawa 240-0193, Japan (E-mail:iwasemin@koryu.w01.boken.ac.jp, Tel:81-468-58-1571, Fax:81-468-58-1544)
FEATURES	Location/Qualifiers
SOURCE	1..7425 /organism="Sus scrofa" /mol_type="genomic DNA" /db_xref="taxon:9823" /sex="male" 1491..7425 /gene="AMELX" join(1491..1548,1140..3151) /gene="AMELX" join(3152..3205,5316..5363,6619..6660,6717..6761, 7033..7425) /gene="AMELX" /codon_start=1 /product="amelogenin" /protein_id="BAC66110.1" /db_xref="gi:29126035" /translation="MGTVILPACLGLAFAFNPRLPRHNGRYINPSYEDLYLAIRID VPAQVPLPGQSHMTPTNHPMLPLPAOOPFOPOVDPPHLOPDPBP MHPIQLPQPPLPFPMFMQSLLPDLPLEAWPAT"
gene	1491..7425 /gene="AMELX" join(1491..1548,1140..3151) /gene="AMELX" join(3152..3205,5316..5363,6619..6660,6717..6761, 7033..7425) /gene="AMELX" /codon_start=1 /product="amelogenin" /protein_id="BAC66110.1" /db_xref="gi:29126035" /translation="MGTVILPACLGLAFAFNPRLPRHNGRYINPSYEDLYLAIRID VPAQVPLPGQSHMTPTNHPMLPLPAOOPFOPOVDPPHLOPDPBP MHPIQLPQPPLPFPMFMQSLLPDLPLEAWPAT"
5'UTR	join(3152..3205,5316..5363,6619..6660,6717..6761, 7033..7425) /gene="AMELX" /codon_start=1 /product="amelogenin" /protein_id="BAC66110.1" /db_xref="gi:29126035" /translation="MGTVILPACLGLAFAFNPRLPRHNGRYINPSYEDLYLAIRID VPAQVPLPGQSHMTPTNHPMLPLPAOOPFOPOVDPPHLOPDPBP MHPIQLPQPPLPFPMFMQSLLPDLPLEAWPAT"
CDS	join(3152..3205,5316..5363,6619..6660,6717..6761, 7033..7425) /gene="AMELX" /codon_start=1 /product="amelogenin" /protein_id="BAC66110.1" /db_xref="gi:29126035" /translation="MGTVILPACLGLAFAFNPRLPRHNGRYINPSYEDLYLAIRID VPAQVPLPGQSHMTPTNHPMLPLPAOOPFOPOVDPPHLOPDPBP MHPIQLPQPPLPFPMFMQSLLPDLPLEAWPAT"
ORIGIN	
Query Match	61.5%; Score 302.8; DB 4; Length 7425;
Best Local Similarity	79.8%; Pred. No. 2.5e-75;
Matches 399, Conservative	0; Mismatches 87; Indels 14; Gaps 3;
Oy	1 ATGGGGAAGCTGATTTGTTCCTCCGCTCGGAGGAGAGCCTTGCAATGCCGTGAGT 60
Dd	3152 ATGGGGAAGCTGATTTGTTCCTCCGCTCGGAGGAGAGCCTTGCAATGCCGTGAGT 3211
Oy	61 AAAATACCCC-----TCATTAATTCAATTTCAACAAGCTTGAAATAAAGTCTGCC 114
Dd	3212 AAAGAACCCCCCTGCTAGTACATGTCCTCAATGTCTACTAGCTTGAAAAAATCTACC 3271
Oy	115 CACAGTTGTTAACTTTAGGGTTTAAGAACAAGTCAAGATGCATGTCTCTCAATGTCTCT 174
Dd	3272 CCGAGTTGATTCACGCTTAGGGATGAAAACAGTAGAAGATCATCTTCTCAAATGACTCA 3331
Oy	175 GTGTTTAGAANAACCTTGGAAGAAGCTTGTTAATAAAAAAAAAATATATATCCCGAGATGCC 234
Dd	3332 GTGTTTAGAANAACCTTGGAAGAAGCTTGTTAATAAAAAAAAAAAGATTCCTCCAGATGCC 3391
Oy	235 ACCCAAGA--CTGATTCAGTAGAGCAGAGATGGGGGAGTCCCGAGACTCTGATTTTA 292
Dd	3392 GCCAAGATTCGATTCAGTAGATAGTGGGGTGG-----GCCATGACTCTGATTTTA 3445
Oy	293 ACAAGCACTCAGAGATTCGTGTGAGACAATTAACTTGTAATATCATGCGCCATCTCT 352
Dd	3446 ACAAGCACCCAGAGAGATCTCTTGAAACAATTAGCTTGTAAGTATCATCACTGTCTCT 3505
Oy	353 AGATGAGAGAACTTTTGAAGAGGACCTTGTAAGGCTCCAGAGAAAGTCTGGAACAG 412
Dd	3506 AGAAGTGGGAACCTTTGGAAGTGAACCTTTGAGAGGCTCCAGAGAAATCTGAAGACG 3565

OY	413	CTTAGGCAAAATACATAAAAAATGCCAATTTCCTTAACAACCATTTCAGACACTGCC	472
Db	3566	CTTAGGCAAAAATAGACAAAAATGTACGTTTCTCTAAAAATCAAATTCCTCGCAAGTGTC	36255
OY	473	AACCTCTCTCCTGCCTCCCA	492
Db	3626	AAATATCTCTTGCTCTCCA	3645
RESULT 4			
AY040206			
LOCUS	AY040206	8810 bp	DNA linear PRI 10-APR-2002
DEFINITION	Homo sapiens amelogenin precursor (AMELX) gene, complete cds.		
ACCESSION	AY040206		
VERSION	AY040206.1	GI:15028582	
KEYWORDS	.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 8810) Hart,P.S., Hart,T.C., Simer,J.P. and Wright,J.T. A nomenclature for X-linked amelogenesis imperfecta Arch. Oral Biol. 47 (4), 255-260 (2002) 21920287 MEDLINE PUBMED 11922868		
REFERENCE	2 (bases 1 to 8810) Hart,S., Hart,T.C. and Wright,T. Direct Submission Submitted (14-JUN-2001) Human Genetics, University of Pittsburgh, 3550 Terrace St., 572A Scalf Hall, Pittsburch, PA 15090, USA		
AUTHORS			
TITLE			
JOURNAL			
FEATURES			
source	1..8810 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /chromosome="X" /map="XP22.3-P22.1"		
gene	1..8791 /gene="AMELX"		
promoter	1..1416 /gene="AMELX"		
TATA_signal	1386..1391 /gene="AMELX"		
mRNA	join(1417..1472,2781..2846,4783..4830,6240..6281,6552..6977,8611..8791). /product="amelogenin precursor"		
exon	1417..1472 /gene="AMELX"		
exon	/number=1 2781..2846 /gene="AMELX"		
CDS	/number=2 join(2793..2846,4783..4830,6240..6281,6552..6977,8611..8616) /gene="AMELX"		
sig_peptide	/codon_start=1 /product="amelogenin precursor" /protein_id="AAK77213.1" /db_xref="GI:15028583" /translation="MGTWLFACILGAFAFMPDPHPDGHGYINFSEVLTPLKYQOS IRPRYSGVYBPMGSMLHNQIIPVLSCDNRPETPLQBNHHIPVLRADQVRIPQDMRV VRSQNSTPIQNHOVPNPPRAQORPQOPQRQDNORMOFRVNHRPRLRRDRFLRR MFEPQSRPRLPDLLAMPSIDTKREVD"		
exon	2793..2840 /gene="AMELX"		
exon	4783..4830 /gene="AMELX"		
exon	/number=3 6107..6148 /gene="AMELX"		

exon	/number=4	6240..6281	/gene="AMELX"
exon	/number=5 <td>6552..6977 <td>/gene="AMELX"</td> </td>	6552..6977 <td>/gene="AMELX"</td>	/gene="AMELX"
exon	/number=6 <td>8611..8791 <td>/gene="AMELX"</td> </td>	8611..8791 <td>/gene="AMELX"</td>	/gene="AMELX"
polYA_signal	/number=7	8785..8791	/gene="AMELX"
ORIGIN			
Query Match	55.8%	Score 274.6; DB 9;	Length 8810;
Best Local Similarity	79.0%	Pred. No. 2.8e-67;	
Matches 394;	Conservative 0;	Mismatches 89;	Indels 16; Gaps 5
QY	1	ATGGGGACCTGGATTGTTGTTGCTGCTCCTCGGGAGCAGCTTTGCTATGCGCCGTAGT	60
Db	2793	ATGGGGACCTGGATTGTTGTTGCTGCTCCTCGGGAGCAGCTTTGCTATGCGCCGTAGT	2855
QY	61	AAAAATACCCCT-----GCATATATTTCAATTTCACAAGCTTGGAAAAATGAAGTGTGCC	114
Db	2853	AAAAACCCCTTGGCATAGTCAAGTGTGCTCAATTTCAAACTTGGACATTAATATCTGCTC	2912
QY	115	CACAGTTGGTAAACTTTAGGGTTTAAACAGTACAAAGTACAGTCTCTCAATGTCTCT	174
Db	2913	ATATGTTGGTGAAGAA--TTAGGGGTTTAAACAGTATGAGATGATGTCTTCATATGTCTCT	2970
QY	175	GTGTTTAAAGAAACACTGGGAAGAGCTGTTATTAATTAATTAATTAATTCAGATGCTCC	234
Db	2971	GGATTGAAGAAACACTT--CAGAGCTTGTTTTAAAGGTATATTTCTAAATGCCGTAC	3029
QY	235	ACCCAGAAGTGAATTCAGTAGACAGAGTGGGGGAGTGGCCAGACTCTGCATTTTAC	294
Db	3030	CAAAATATCTGATTTGGTATAGCTGTGGGGGGG-----GCCAGAGCTCTGCATTTTAT	3083
QY	295	AGCACCTTCAGAGAGTTCTGTGAGACAAATTACTTTGTAATATATCATGCCCATCTCTAG	354
Db	3084	AGCACCCCGAGAGATTCGTGTGAACTGTAGCTGTATTAATATCACCAACCATCTCTAG	3143
QY	355	ATGAGAGAACTTTTAAAGAGGAGCCCTTGAAAGGCTCCAGAGAAAGTCTGGAACGCT	414
Db	3144	ATGAGAGAACTTTTGAAGAGGAGCCCTTGAAAGGCTCCAGAGAAAGTCTTAAACGACT	3203
QY	415	TAGG--CAAAATCTACAAAAATGCAATTTTCTTAAACCCAAATTTCTTAAAGAGTGTCCA	473
Db	3204	TTGGCAAAATATTAACAGAGATGCCAGTTTGTCTTAAACCCAAATTTCTCTCAAGATTTCCA	3263
QY	474	ACTCTCTCTCGGCCCTTCCA 492	
Db	3264	AATCTCTCTCTGCCCTTCCA 3282	
RESULT 5			
AL357130/c	158142 bp	DNA	linear
LOCUS	AL357130	158142 bp	DNA linear
DEFINITION	Homo sapiens chromosome X clone RP13-169E15, 4 unordered pieces.		
ACCESSION	AL357130.3	GI:9863914	
VERSION	AL357130.3	GI:9863914	
KEYWORDS	HTG; HTGS; PHASE1; HTGS_CANCELED.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	McLay, K.		
JOURNAL	Direct Submission		
COMMENT	Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humgenet@sanger.ac.uk		
	requests: clonerequest@sanger.ac.uk		
	On Aug. 21, 2000 this sequence version replaced gi:9214076.		

```

----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: Bb169E15
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; 108752; 100% of reads
Chemistry: dye-terminator Big Dye; 100% of reads
Consensus quality: 155448 bases at least Q40
Consensus quality: 156466 bases at least Q30
Consensus quality: 157038 bases at least Q20
Insert size: 157842; sum-of-contigs
Insert size: 160705; 33.4% error; agarose-fp
Quality coverage: 4.56x in Q20 bases; sum-of-contigs quality
coverage: 4.56x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 68897: contig of 68897 bp in length
* 68898 68997: gap of 100 bp
* 68998 122842: contig of 53845 bp in length
* 122843 122943: gap of 100 bp
* 122943 129584: contig of 6642 bp in length
* 129585 129685: gap of 100 bp
* 129685 158142: contig of 28458 bp in length.
*
Location/Qualifiers
1. 158142
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="X"
/clone="RP13-169E15"
/clone_1b="RPC1-13.1"
1. 68897
/note="assembly fragment: 01108
fragment_chain:1
clone_end:17
vector_side:left"
68998. 122842
/note="assembly fragment: 00101
fragment_chain:1"
122943. 129584
/note="assembly fragment: 00361
fragment_chain:1"
129685. 158142
/note="assembly fragment: 01285
fragment_chain:1
clone_end:SP6
vector_side:right"
-----
ORIGIN
Query Match 55.8%; Score 274.6; DB 2; Length 158142;
Best Local Similarity 79.0%; Pred. No. 2,9e-67;
Matches 394; Conservative 0; Mismatches 89; Indels 16; Gaps 5;
QY 1 ATGGGACCGTGATTTTGTGCTGCTCCCTCGGAGCGACGCTTGGCTATGCCCCGTAGT 60
Db 44325 ATGGGACCGTGATTTTATTTATTTGCTCCCTCGGAGCGACTTTTGGCCATGCGTGTAGT 44266
QY AAAATACCCCT-----GCATATATTTCAATTTACAAAGCTTGAATAAAGTCGCGCC 114
Db 44265 AAAAAGCCCTTGATTAAGTCAAGTGTCCCAATTTTCAACAACCTTGACATTAATAATCTGCTC 44206
QY 115 CACAGTTTGTAACCTTTAAGGTTTAAAGACGTAACAAGATCAAGATGTCTCAATGTCTCT 174

```







QY 355 ATGAGGAACTTTAGAGGACCTTG--AAAGCCTCAGAGAAAGTCTGACAG 412  
 DB 1868 ATGAGGAAAGTCTCAGAGGACCTTGAGAAAGGCTCTTGAGAAAGCTTAATCAG 1927  
 QY 413 C-TTAGGCAATATACCAAAATGCCAATTTCTCTAAACCAATTTCTACAGATGC 471  
 DB 1928 CATTAGGCAATATATACCAAAATGCCAATTTCTCTAAACCAATTTCTC--CAAGTTTC 1985  
 QY 472 CAATCTCTTCTGCTCCCA 492  
 DB 1986 CAAATTTCTTCACGCTCCCA 2006  
 RESULT 9  
 LOCUS AB091783 6442 bp DNA linear PRI 02-MAY-2003  
 DEFINITION Saimiri sciureus AMELX gene for amelogenin, partial cds.  
 ACCESSION AB091783  
 VERSION AB091783.1 GI:29126019  
 KEYWORDS  
 SOURCE Saimiri sciureus (common squirrel monkey)  
 ORGANISM Saimiri sciureus  
 Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Cebinae;  
 Saimiri.  
 REFERENCE 1  
 AUTHORS Iwase, M., Saita, Y., Hirai, Y., Hirai, H., Imai, H. and Takahata, N.  
 TITLE From the Cover: The amelogenin loci span an ancient pseudautosomal  
 boundary in diverse mammalian species  
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 100 (9), 5258-5263 (2003)  
 MEDLINE 22608569  
 PUBMED 12672962  
 REFERENCE 2 (bases 1 to 6442)  
 AUTHORS Iwase, M., Saita, Y. and Takahata, N.  
 TITLE Direct Submission  
 JOURNAL Submitted (19-SEP-2002) Mineyo Iwase, Graduate University for  
 Advanced Studies (Sokendai), Department of Biosystems Science;  
 Shonan Kokusaiimura, Hayama, Kanagawa 240-0193, Japan  
 (E-mail: iwase@imn.koryu.w01.soken.ac.jp, Tel:81-468-58-1571,  
 Fax:81-468-58-1544)  
 FEATURES  
 source Location/Qualifiers  
 1..6442  
 /organism="Saimiri sciureus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9521"  
 /sex="male"  
 936..6442  
 /gene="AMELX"  
 /join(936..991,2298..2309)  
 /gene="AMELX"  
 /join(2310..2363,4289..4336,5592..5633,5725..5769,  
 6041..>6442)  
 /gene="AMELX"  
 /codon\_start=1  
 /product="amelogenin"  
 /protein\_id="BAC66103.1"  
 /db\_xref="GI:29126020"  
 /translation="MGTWILFACILGAAPAMP.PHPGHPGYNSFENSHQATNID  
 RNLVLTPLKYSMTLRAPRSYSGYPRMGSHLHQTIPLSQHPPTHTLQPHHNPV  
 VPAQPVVFOQPMKPVPGQHSMTPTQHQNPFPQAQPPQPTQVQPHQPMQPPQPP  
 VHPMQPLPPQPLPPMFPMPQPLPPMLPDLPLEAMPAT"  
 ORIGIN  
 Query Match 52.2%; Score 256.6; DB 9; Length 6442;  
 Best Local Similarity 79.3%; Pred. No. 3.8e-62;  
 Matches 394; Conservative 0; Mismatches 89; Indels 14; Gaps 7;  
 QY 1 ATGAGGAACTGATTTGTTGCTGCTCTGAGAGAGAGCTTGTATAGCCGTGAGT 60  
 DB 2310 ATGAGGAACTGATTTGTTGCTGCTCTGAGAGAGAGCTTGTATAGCCGTGAGT 2369  
 QY 61 AAAA---TACCCCTGATATATATCAATTTCAAGAGCTTGAAATAAAGTCTGCCCA 116

DB 2370 AAAACCCCTTGATAGTACAGTCTCAATTTCTATTAAGATTAGCAATAAAGTCTGCCCA 2429  
 QY 117 CAGTTGTAACCTTTAGGGTTTAAAGACAGTACAGATGATGCTCTCAATGCTCTGT 176  
 DB 2430 TGGTTGGTGA--GTAGGGCTTTAAACAGTATGAGTAAAGTCTTCAACATTTCTGG 2488  
 QY 177 GTTTAAGAAACCTTGGAAGAGCTTGTATTAATAAATAATATATTCCTCCAGATGCTCCAC 236  
 DB 2489 GTTGAAGAAACACTT-CAGGAGCTGTTT--AAAAAGTAAATTTCAAAATGCTCCAG 2545  
 QY 237 CCAAGACTGATTCAGTAGAGCAGAGTGGGGAGAGTCCAGAGCTCTGATTTTAAACA 296  
 DB 2546 CAAA---GATTCGATTTTGTACAGCTGGGGTGGGGCCAGAGCTCTGATTTTAAACA 2601  
 QY 297 GCACCTCAGAGATTTCTGTGAGACATTAATTAATATATATCCATCTCTAGAT 356  
 DB 2602 GTACCAACAGAGATTTCTGTGAACCATAGCTGTGATATGTATACCAACCATCTCTAGAT 2661  
 QY 357 GGAGGAACTTTTAGAGAGGACCTTGAAAGGCTCCAGAGAAAGTCTCGAAGCAGC--TT 415  
 DB 2662 GGAGGAACTTTTAGAGAGGACCTTGAAAGGCTCTTAGAGAAAGTCTTCAACAGCTTT 2721  
 QY 416 AGGCAATATCTACAAAATGCCAATTTCTCTTAAACCAATTTCTTACAGATGCTCAAC 475  
 DB 2722 GGGCAAAATTTACAGAGATGCCAGTTTGTCTTAAA--CCAATTCCTCAAGTTTCCAAA 2780  
 QY 476 TCTCTTCTGCTCCCTCCA 492  
 DB 2781 TATCTTCTGCTCCCTCCA 2797

RESULT 10  
 LOCUS AB091794 5591 bp DNA linear MAM 02-MAY-2003  
 DEFINITION Equus caballus AMELX gene for amelogenin, partial cds.  
 ACCESSION AB091794  
 VERSION AB091794.1 GI:29126040  
 KEYWORDS  
 SOURCE Equus caballus (horse)  
 ORGANISM Equus caballus  
 Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Perissodactyla; Equidae; Equus.  
 REFERENCE 1  
 AUTHORS Iwase, M., Saita, Y., Hirai, Y., Hirai, H., Imai, H. and Takahata, N.  
 TITLE From the Cover: The amelogenin loci span an ancient pseudautosomal  
 boundary in diverse mammalian species  
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 100 (9), 5258-5263 (2003)  
 MEDLINE 22608569  
 PUBMED 12672962  
 REFERENCE 2 (bases 1 to 5591)  
 AUTHORS Iwase, M., Saita, Y. and Takahata, N.  
 TITLE Direct Submission  
 JOURNAL Submitted (19-SEP-2002) Mineyo Iwase, Graduate University for  
 Advanced Studies (Sokendai), Department of Biosystems Science;  
 Shonan Kokusaiimura, Hayama, Kanagawa 240-0193, Japan  
 (E-mail: iwase@imn.koryu.w01.soken.ac.jp, Tel:81-468-58-1571,  
 Fax:81-468-58-1544)  
 FEATURES  
 source Location/Qualifiers  
 1..5591  
 /organism="Equus caballus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9796"  
 /sex="male"  
 130..5591  
 /gene="AMELX"  
 /join(130..180,1483..1494)  
 /gene="AMELX"  
 /join(1495..1548,3470..3517,4907..4915,5235..>5591)  
 /gene="AMELX"  
 /codon\_start=1  
 /product="amelogenin"  
 /protein\_id="BAC66113.1"

## ORIGIN

/db xref="GI:29126041"  
/translation="MGWTLFACLVGRTAAMPBPHGHPYINFSYKSYGVE  
PMGKMLHNOIIPUTQONPSNHLQPHNHIPMPAQOPVVSQOPALIPQCHSITPSQ  
HHOBNLPVPQOPPOPHPIOPQPHMHPLOPLPQAPLPPLPPLAASPHAS"

Query Match 50.6%; Score 248.8; DB 4; Length 5591;  
Best Local Similarity 72.4%; Pred. No. 6.4e-60;  
Matches 356; Conservative 0; Mismatches 122; Indels 14; Gaps 2;

QY 1 ATGGGACCTGATTTTGTGCTGCTGCTGAGACAGCTTGTATGCCGTAGT 60  
DB 1495 ATGGGACCTGATTTTGTGCTGCTGCTGAGACAGCTTGTATGCCGTAGT 1554  
QY 61 AAAATACCTGCTAATATTCATTTCAAGCTTGAATAAAGTCTGCCACAGT 120  
DB 1555 ATAAATC-----ACATCTCTATTTCACAGCTTGAATAAAGTCTGCCACAGT 1606  
QY 121 TGGTAACTTTAGGGTTTAAAGACATCAAGATGATGATGATGATGATGATGAT 180  
DB 1607 TGGTAACTTTAGGGTTTAAATCAATCAAGATGATGATGATGATGATGATGAT 1666  
QY 181 AAGAACCTTGAAGAGCTTGTATTAATAAATAATATTCAGATGCTCCACCAA 240  
DB 1667 AAGATTAATTTGAAGAGCTTGTATTAATAAATAATATTCAGATGCTCCACCAA 1726  
QY 241 GACTGATTCAGTACGACGAGAGTGGGGAGTCCCGACACTGTCAATTTACACAGAC 300  
DB 1727 TTTGATTTGAATACGCTGAGAGAGG-----ACCGAGACTCTGTCTTTTAATTAACCC 1780  
QY 301 CTCAGAGATTTCTGAGACAAATTAATTTGATTAATTCATGCTCCATCTGATGAGAG 360  
DB 1781 CCAGAGATTTCTGAGACAAATTAATTTGATTAATTCATGCTCCATCTGAGAGAG 1840  
QY 361 GAAACTTTTGAAGAGAGCTTGAAGAGCTCCAGAGAAAGTCTCCAGAGCTTGAAGCA 420  
DB 1841 GAAAGTTTGAAGAGAGCTTGAAGAGAGCTCCAGAGAAAGTCTTGAAGAGCTTGAAGCA 1900  
QY 421 AATATCTAATAAATGCAATTTTCTTAAACCAATTTCTTAAAGAGTCTCAACTCTCT 480  
DB 1901 AACCTGACAAACATGATGCTCTTCAACAGCAATTTCTCAACAGTGTATTAATTTCT 1960  
QY 481 TCCGCTCCCA 492  
DB 1961 CCTGACTCTCCA 1972

## RESULT 11

AB091787 5562 bp DNA linear PRI 02-MAY-2003  
LOCUS AB091787  
DEFINITION Oclemur garnettii AMELX gene for amelogenin, partial cds.  
ACCESSION AB091787.1 GI:29126027  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

Oclemur garnettii (small-eared galago)  
Oclemur garnettii  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Strepsirrhini; Galagonidae; Oclemur.

1 Iwase, M., Satta, Y., Hirai, Y., Hirai, H., Imai, H. and Takahata, N.  
From the Cover: The amelogenin loci span an ancient pseudautosomal  
boundary in diverse mammalian species  
Proc. Natl. Acad. Sci. U.S.A. 100 (9), 5258-5263 (2003)

## REFERENCE

JOURNAL MEDLINE  
PUBMED 22608569  
2 (bases 1 to 5562)  
Iwase, M., Satta, Y. and Takahata, N.  
Direct Submission  
Submitted (19-SEP-2002) Mineyo Iwase, Graduate University for  
Advanced Studies (Sokeidai), Department of Biosystems Science;  
Shonan Kokueiimura, Hayama, Kanagawa 240-0193, Japan  
(E-mail: iwase@imn.koryu.w01.soken.ac.jp, Tel: 81-468-58-1571,  
Fax: 81-468-58-1544)

## FEATURES

## Source

1..5562  
/organism="Oclemur garnettii"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:30611"  
/sex="male"  
267..5562  
/gene="AMELX"  
join(267..321,1401..1412)  
/gene="AMELX"  
join(1413..1466,3385..3432,4699..4740,4832..4876,  
5125..5562)  
/gene="AMELX"  
/codon\_start=1  
/product="amelogenin"  
/protein\_id="BAC6107.1"  
/db\_xref="GI:29126028"  
/translation="MGWTLFACLVGRTAAMPBPHGHPYINFSYKSYGVE  
PMGKMLHNOIIPUTQONPSNHLQPHNHIPMPAQOPVVSQOPALIPQCHSITPSQ  
HHOBNLPVPQOPPOPHPIOPQPHMHPLOPLPQAPLPPLPPLAASPHAS"

## gene

5'UTR  
CDS

## ORIGIN

Query Match 49.8%; Score 245.2; DB 9; Length 5562;  
Best Local Similarity 74.9%; Pred. No. 6.8e-59;  
Matches 379; Conservative 0; Mismatches 108; Indels 19; Gaps 5;

QY 1 ATGGGACCTGATTTTGTGCTGCTGCTGAGACAGCTTGTATGCCGTAGT 60  
DB 1413 ATGGGACCTGATTTTGTGCTGCTGCTGAGACAGCTTGTATGCCGTAGT 1472  
QY 61 AAAATACCTGCTAATATTCATTTCAAGCTTGAATAAAGTCTGCCACAGT 114  
DB 1473 AAAACACCTTGAAGAGCTTGTATTAATAAATAATATTCAGATGCTCCACCAA 1532  
QY 115 CACAGTGTAACTTTAGGGTTTAAAGACATCAAGATGATGATGATGATGATGAT 174  
DB 1533 AACAGTGTAACTTTAGGGTTTAAACAGATGATGATGATGATGATGATGAT 1592  
QY 175 GTGTTAAGAACACTTGAAGAGCTTGTATTAATAAATAATATTCAGATGCTCC 234  
DB 1593 CTCTGTTGAAGAACACTTGAAGAGATTA-CTTAAAGTATGATCTCAAGAGCTCC 1651  
QY 235 ACCAAGACTGATTCAGTACGACGAGAGTGGGGAGTCCCGACACTGTCAATTTAAC 294  
DB 1652 ACCAAA-----GATTCGATTTTGAAGACGAGAGCAAGGCCGCTCGAGTTTTCAC 1707  
QY 295 AAGCAGCTGAGAGATTTCT-GTGAGACAAATTAATTTGATTAATGATGATGATGAT 353  
DB 1708 AAGCAGCTTGAAGAGATTTCTGATTAACCAATTTGATTAATGATGATGATGAT 1767  
QY 354 GATGAGAGAACTTTTGAAGAGAGCTTGAAGAGCTCCAGAGAAAGTCTGAAACAGC 413  
DB 1768 GATGAGAGAACTTTTGAAGAGAGCTTGAAGAGCTCCAGAGAAAGTCTGAAACAGC 1827  
QY 414 TTAGCGAAA-----TACTACAAAATGCAATTTTCTTAAACCAATTTCTAAGCA 466  
DB 1828 CTGAGCAATATACCTGATTAACAAATCAAGATTTTCTTAAAGCCAGGTTCTTTCAA 1887  
QY 467 GTGTCAACTCTTCTGCTGCCCA 492  
DB 1888 GTTTCAAATCTTCCATCTCTCA 1913

## RESULT 12

DB3730 1935 bp DNA linear PRI 10-JUL-2001  
LOCUS DB3730  
DEFINITION Homo sapiens AMGX gene for amelogenin, intron 2.  
ACCESSION DB3730.1 GI:5263179  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM



REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
FEATURES  
source

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1  
Sekiuchi, H., Minaguchi, K., Machida, Y. and Yakushiji, M.  
PCR detection of the human amelogenin gene and its application to the diagnosis of amelogenesis imperfecta  
Bull. Tokyo Dent. Coll. 39 (4), 275-285 (1998)  
99234629  
10218009  
2 (bases 1 to 1935)  
Sekiuchi, H.  
Direct Submission  
Submitted (29-FEB-1996) Hiroshi Sekiuchi, Tokyo Dental College, Pediatric Dentistry, 1-2-2 Masago, Mihama-ku, Chiba, Chiba 261, Japan (Tel:043-270-3945, Fax:043-279-2052)  
Location/Qualifiers  
1. .1935  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="X"  
1. .1935  
/gene="AMGX"  
1. .1935  
/gene="AMGX"  
/note="amelogenin  
this nucleotide sequence shows an intron between exon2 and exon3 within human amelogenin gene on the X chromosome"  
/number=2

ORIGIN

Query Match 44.6%; Score 219.6; DB 9; Length 1935;  
Best Local Similarity 78.5%; Pred. No. 1.4e-51;  
Matches 328; Conservative 0; Mismatches 79; Indels 11; Gaps 5;

Y 76 AATATCAATTTCAAGCTTGAATATAAGTCTGCCACAGTTGTAACCTTTAGG 135  
D 28 AGTCTCCATTTCCAACTTGGACATTAATTTGCCCATAGTGTGTA--TTAGG 85  
Y 136 TTAAAGACGTACAAATGATCTCTCAATGCTCTGTGTTTAAAGAACTTTGAA 195  
D 86 TTAAACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 144  
Y 196 GAGCTTTTAAATATAATATATATCCAGATGCCCTCCACCAAGCTGATTCAGAGA 255  
D 145 GAGCTTTTAAATATAATATATATCCAGATGCCCTCCACCAAGCTGATTCAGAGA 204  
Y 256 GCAAGAGTGGGGGAGTGGCCAGGACTCTGCAATTTTAAAGCACTCAGAGATTCGT 315  
D 205 GCTGGGGGGG-----GCCAGAGACTCTGCAATTTTAAAGCACTCAGAGATTCGT 258  
Y 316 GGAAGCAATTAATCTTGAATATATATGATGATGATGATGATGATGATGATGAT 375  
D 259 TGGAACTGTTAGCTTTGAAATATATACCAATCTTGAATGAGAGCTTTTGAAG 318  
Y 376 GACCTTGAAGAGCTTCAAGAGAGCTTCAAGCTTGAAG--CAATTAATAAAT 434  
D 319 GACCTTGAAG--GTCTCAAGAGAGCTTCAAGCTTGAAGCAATTAATAAAGAT 377  
Y 435 GCCAATTTTCTTAAAGCCCAATTTTCAAGAGAGCTTCAAGCTTCAAGCTTCAAG 492  
D 378 GCCAATTTTCTTAAAGCCCAATTTTCTTCAAGATTCCTTCTGCTGCTGCA 435

RESULT 13  
AB091790 6264 bp DNA linear MAM 02-MAY-2003  
LOCUS Bos taurus AMELY gene for amelogenin, partial cds.  
ACCESSION AB091790  
VERSION AB091790.1 GI:29126032  
KEYWORDS  
SOURCE Bos taurus (cow)  
ORGANISM Bos taurus

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
FEATURES  
source

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.  
1  
Iwase, M., Satta, Y., Hirai, Y., Hirai, H., Imai, H. and Takahata, N.  
From the Cover: The amelogenin loci span an ancient pseudautosomal boundary in diverse mammalian species  
Proc. Natl. Acad. Sci. U.S.A. 100 (9), 5258-5263 (2003)  
22608569  
12672962  
2 (bases 1 to 6264)  
Iwase, M., Satta, Y. and Takahata, N.  
Direct Submission  
Submitted (19-SEP-2002) Mineo Iwase, Graduate University for Advanced Studies (Sokendai), Department of Biosystems Science, Shonan Kokusaiimura, Hayama, Kanagawa 240-0193, Japan (E-mail: iwase@nhe.koryu.w01.soken.ac.jp, Tel:81-468-58-1571, Fax:81-468-58-1544)  
Location/Qualifiers  
1. .6264  
/organism="Bos taurus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9913"  
/sex="male"  
114. .6264  
/gene="AMELY"  
join(114. .169,1585. .1596)  
/gene="AMELY"  
join(1597. .1659,3515. .3562,5080. .5124,5863. .>6264)  
/gene="AMELY"  
/codon\_start=1  
/product="amelogenin"  
/protein\_id="BA066109.1"  
/db\_xref="GI:29126032"  
/translation="MGTWLFACLLGAGVSMVSIIPHPHPGVINSEYVUTPLKMYNMRLRPVPSGYEYVGVGMHRIIPVVSQSPQNALQPHHNPMVAQDPVPOQPMVPGHSMPIQHHQPNLPPLPAQSFQPPPIOPPPPPVHPHPIORLPPOPPIPPPPMPQPLPVLPLPLPAMPAT"

ORIGIN

Query Match 33.8%; Score 166.2; DB 4; Length 6264;  
Best Local Similarity 68.0%; Pred. No. 2.4e-36;  
Matches 338; Conservative 0; Mismatches 118; Indels 41; Gaps 6;

Y 1 AATGAGACCTGGAATTTGTTGCTGCTGCTGCGGAGAGCCCTTGTAAGCCGAGT 60  
D 1597 AATGAGACCTGGAATTTGTTGCTGCTGCTGCGGAGAGCCCTTGTAAGCCGAGT 1656  
Y 61 AATATACCCCTGCATATATTCATTTCAAGCTTGAATATAAGTCTGCCACAGT 120  
D 1657 AATATAC-----AATAGCTAATTTTCCAAATTTGGAATAAATATGCAACAAT 1708  
Y 121 TGGTAACTTTAGGTTTAAAGACATGATGATGATGATGATGATGATGATGATGAT 180  
D 1709 AGGTAACTTTCAAGTGTGTT-----AGATTAAGATCAATATGCTTAAATGCTGTG 1763  
Y 181 AAGAAACCTGGAAGAGCTGTTAT--AAAAAATATATATCCAGATGCTGCAC 237  
D 1764 AAGATCAATTTGAAGAGCTGTTATTAATAAATAAAGATCCCAAGATTTTCTGCT 1823  
Y 238 CAAGA--CTGATTCATGAGAGAGAGTGGGGGAGTGGCCAGAGCTTGCAATTTTAA 295  
D 1824 AAGATTTGATTCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1861  
Y 296 AGCACTCAGAGATTTCTGAGAGCAATTAATTTGAATATATATATATATATATAT 355  
D 1862 AATATGAGAGAGATTCATGATGATGATGATGATGATGATGATGATGATGATGAT 1921  
Y 356 TGGAGAACTTTTAAAGGAGCCCTGGAAGGCTCCAGAGAAAGTCTGCAAGCTT 415  
D 1922 TGAAGGAA--TTTGGAGGAGCTTCAAGAGCTGGAAGATTAATTTAATCAATGT 1980  
Y 416 AGCAAAATCTACAAATATGCAATTTTCTTAAATCCCAATTTTCAAGAGTGTCAAC 475



```

Db      1981 TAGTTAAATACAAATATCTACTTTACTTAAATCCAAATTTCTTACAAAGTTATTA 2040
Qy      476 TCTCTCTCGGCTCCCA 492
Db      2041 TATTTCTCTGCCCCCA 2057

RESULT 14
AF294397      9384 bp   DNA   linear   ROD 03-SEP-2000
DEFINITION   Mus musculus amelogenin gene, promoter and partial cds.
ACCESSION    AF294397
VERSION      AF294397.1   GI:9965405
KEYWORDS
SOURCE       Mus musculus (house mouse)
ORGANISM     Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE    Sneed,M.L., Paine,M.L., Luo,W., Zhu,D.H., Yoshida,B., Lei,Y.P.,
AUTHORS      Paine,C.T., Chen,L.S., Burslein,J.M., Jilpukdeebudintira,S.,
              White,S.N. and Birling,P. Jr.
              Transgene animal model for protein expression and accumulation into
              forming enamel
              Connect. Tissue Res. 38 (1-4), 279-286 (1998)
JOURNAL      20515040
MEDLINE      11063035
REFERENCE    2 (bases 1 to 9384)
AUTHORS      Sneed,M.L., Zhu,D.-H., Lei,Y.-P. and Paine,M.L.
TITLE        Direct Submission
JOURNAL      Submitted (08-AUG-2000) Dentistry, University of Southern
              California, 2250 Alcazar Street, CSA103, Los Angeles, CA 90033, USA
FEATURES
    source
        1. .9384
           /organism="Mus musculus"
           /mol_type="genomic DNA"
           /db_xref="taxon:10090"
           /chromosome="X"
           1..8546
           promoter
           TATA_signal
           mRNA
           CDS
               8547..7194
               /product="amelogenin"
               8559..>8612
               /note="enamel matrix protein"
               /codon_start=1
               /product="amelogenin"
               /protein_id="AAG10078.1"
               /db_xref="GI:9965406"
               /translation="MGTWILFACLLGAAPMP"

ORIGIN
Query Match      32.6%; Score 160.6; DB 10; Length 9384;
Best Local Similarity 65.7%; Pred. No. 9,7e-35;
Matches 332; Conservative 0; Mismatches 154; Indels 19; Gaps 6;

Qy      1 ATGGGACCTCGATTTTGTGCTGCTCTCGGAGACACCTTTGCTATGCCCGTAGT 60
Db      8559 ATGGGACCTCGATTTTGTGCTGCTCTCGGAGACACCTTTGCTATGCCCGTAGT 8618

Qy      61 AAAATACCCC-----TGCATATATTTCAATTTTCACAGCTTGGAATTAAGTCTGC-CC 114
Db      8619 AAAACATCATTTACTTACATTTGCAATTCATTAATCTTGGAACTAGAAATCTGCTCC 8678

Qy      115 CACAGTTGTAACCTTTAGGCTTTAAGACAGTACCAAGATCAGATGCTCCAAATGCTCT 174
Db      8679 CATAGTGTGTAATTTTCACTGTGTACACAGTCAAGATCTTATATCCCAAGATCTAA 8738

Qy      175 GTGTTTAAAGAAACCTTGAAGAGCTGTTATTAATAAATAATATTTCCAGATGCTCC 234
Db      8739 TAGGTGTTGAAGAAATATTTTGAAGTTTGTAAAGAA-----ATTCAGAGAAAGCTCC 8794

Qy      235 ACCCAAGAC--TGATTCAGTAGAGACAGAGTGGGGGAGTGCCAGGACTCTGCATTTA 292

```

```

Db      8795 ACCTTAGACTGTATTCAGTTACGTGAACTGGAGACCAAGACTCAAGATTTTAACAGC 8854
Qy      293 ACAAGACCTCAGAGATTTCTGTGAGACAAATTACTGTAAATATCATGCCATCTCT 352
Db      8855 ATTTGACCCCTAGAGATCTTCTCTTAACATTAAGCTGTAAACACATGCCCATCTCT 8914

Qy      353 AGATGAGGAAATCTTTAGAGGAGACCTTGAAAGGCTTCAGAGAAAGTCTCGAAACAG 412
Db      8915 ACAAGTAGTATTCATTTTGAAGGAGACCTTGAAAGGCTTCAGAG--AGTGTACTGTGT 8972

Qy      413 CTAGGCAATATACAAAATAGC-----ATTTCTCTAAAAACCAATTTCTAAGAG 467
Db      8973 TTTAGGCAATATTAACAGCCCCCCCCCAAAAATATCTTAAGCTTAATTACTCAGAG 9032

Qy      468 TGTCAACTCTCTTCTGCCCCCA 492
Db      9033 TTCCAAAATTTCTTCTGCCCCCA 9057

RESULT 15
AL805974      95826 bp   DNA   linear   ROD 29-AUG-2003
LOCUS        AL805974/c
DEFINITION   Mouse DNA sequence from clone Rp23-334F21 on chromosome X, complete
              sequence.
ACCESSION    AL805974
VERSION      AL805974.8   GI:34366495
KEYWORDS
SOURCE       Mus musculus (house mouse)
ORGANISM     Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE    1 (bases 1 to 95826)
AUTHORS      Chapman,J.
TITLE        Direct Submission
JOURNAL      Submitted (28-AUG-2003) Wellcome Trust Sanger Institute, Hinxton,
              Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
              humquerry@sanger.ac.uk
              On Aug 30, 2003 this sequence version replaced gi:25955748.
              Sequence from the Mouse Genome Sequencing Consortium whole genome
              shotgun may have been used to confirm this sequence. Sequence data
              from the whole genome shotgun alone has only been used where it has
              a phred quality of at least 30.
              ----- Genome Center
              Center: Wellcome Trust Sanger Institute
              Center code: SC
              Web site: http://www.sanger.ac.uk
              Contact: humquerry@sanger.ac.uk
              -----

COMMENT
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest, except on the rare
occasion of the clone being a YAC.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMBPP; Information
on the WORMBPP database can be found at
http://www.sanger.ac.uk/projects/C_elegans/wormpep
from the RPI-23 Mouse BAC Library
constructed by the group of Pieter de Jong.
For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBAC3.6.
Location/Qualifiers

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 8, 2005, 11:25:10 : Search time 652.659 Seconds  
(without alignments)  
3957.220 Million cell updates/sec

Title: US-10-754-437-22

Perfect score: 492  
Sequence: 1 atgsgggagccggatctgtc.....aactctctctgcctcca 492

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 413486 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_23Sep04:\*  
1: geneseq1980s:\*  
2: geneseq1990s:\*  
3: geneseq2000s:\*  
4: geneseq2001a:\*  
5: geneseq2001b:\*  
6: geneseq2002a:\*  
7: geneseq2002b:\*  
8: geneseq2003a:\*  
9: geneseq2003b:\*  
10: geneseq2003c:\*  
11: geneseq2003d:\*  
12: geneseq2004a:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	274.6	55.8	259202	12	ADQ18492
2	51.4	10.4	270	12	ADM08035
3	51.4	10.4	549	12	ADM08036
4	51.4	10.4	623	12	ADM08037
5	51.2	10.4	476	3	AAZ50832
6	50.2	10.2	556	10	ADBS9026
7	50.2	10.2	556	10	ADBS3782
8	49.8	10.1	722	6	AA14111
9	49.8	10.1	752	6	AA14110
10	48.2	9.8	318	12	ADM08036
11	48.2	9.8	750	2	AAZ07020
12	48.2	9.8	793	2	AAZ07018
13	46.6	9.5	802	2	AAZ07019
14	46.6	9.5	852	12	ADQ22977
15	45.2	9.2	65	6	ABN28389
16	39.8	8.1	95000	12	ADOS6276
17	39.4	8.0	216215	10	ADFS9167
18	39.2	8.0	97658	8	ABQ83210_3
19	37.4	7.4	5310	8	AAK65863
20	36.2	7.4	595	12	ACH74683
21	36.2	7.4	166910	12	ADN01278

C	22	35.8	7.3	374	2	AAV78614	AAV78614	Staphyloc
C	23	35.8	7.3	742	2	AAV77973	AAV77973	Staphyloc
C	24	35.8	7.3	1468	8	ADA89829	ADA89829	Staphyloc
C	25	35.8	7.3	1665	8	ACF74621	ACF74621	Staphyloc
C	26	35.8	7.3	10758	2	AAV74532	AAV74532	Staphyloc
C	27	35.6	7.2	7536	10	ADE53812	ADE53812	Human pro
C	28	35.2	7.2	25001	10	ADC87256	ADC87256	Human GPC
C	29	34.8	7.1	2282	4	AAH17568	AAH17568	Human CDN
C	30	34.8	7.1	14252	4	AA104590	AA104590	Human rep
C	31	34.8	7.1	14252	4	AA104692	AA104692	Human rep
C	32	34.8	7.1	14252	4	AB197599	AB197599	Human tes
C	33	34.8	7.1	14252	4	AB197513	AB197513	Human tes
C	34	34.8	7.1	16545	6	AB132050	AB132050	Human tes
C	35	34.8	7.1	24200	4	AB121182	AB121182	Human tes
C	36	34.8	7.1	24206	4	AB121180	AB121180	Drosophila
C	37	34.8	7.1	24235	4	AB105026	AB105026	Drosophila
C	38	34.8	7.1	300000	10	ADBS6352	ADBS6352	Human PTP
C	39	34.8	7.1	300001	12	ADQ14076	ADQ14076	Human pro
C	40	34.6	7.0	837	4	AAH73049	AAH73049	Human cer
C	41	34.6	7.0	1878	10	ACC61054	ACC61054	Gene sequ
C	42	34.6	7.0	1878	10	ADK62759	ADK62759	Disease t
C	43	34.6	7.0	2000	8	ADA68928	ADA68928	Arabidops
C	44	34.2	7.0	1668	8	ACA46163	ACA46163	Prokaryot
C	45	34.2	7.0	1668	8	ACA46163	ACA46163	Prokaryot

#### ALIGNMENTS

RESULT 1	ADQ18492	standard; DNA; 259202 BP.
ID	ADQ18492;	
AC	ADQ18492;	
XX	26-AUG-2004 (first entry)	
DT	26-AUG-2004 (first entry)	
DE	Human soft tissue sarcoma-upregulated DNA - SEQ ID 1311.	
XX	Human soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human; de.	
KW	Human soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human; de.	
OS	Homo sapiens.	
XX	26-NOV-2003; 2003MO-US038193.	
PF	26-NOV-2003; 2003MO-US038193.	
XX	26-NOV-2002; 2002JUN-0429739P.	
PR	26-NOV-2002; 2002JUN-0429739P.	
XX	(PROT-) PROTEIN DESIGN LABS INC.	
PA	(PROT-) PROTEIN DESIGN LABS INC.	
XX	Aziz N, Glinesburg WM, Zlotnick A;	
PI	Aziz N, Glinesburg WM, Zlotnick A;	
DR	WPI; 2004-441208/41.	
XX	Early detection of soft tissue sarcoma comprises determining expression of a gene in a first soft tissue sample and a normal soft tissue sample and comparing the gene expression, also useful in treating soft tissue sarcoma.	
PT	Early detection of soft tissue sarcoma comprises determining expression of a gene in a first soft tissue sample and a normal soft tissue sample and comparing the gene expression, also useful in treating soft tissue sarcoma.	
XX	Example 2; SEQ ID NO 1311; 210bp; English.	
PS	Example 2; SEQ ID NO 1311; 210bp; English.	
XX	The invention relates to a novel method for detecting soft tissue sarcoma which comprises obtaining a first soft tissue sample from an individual, and a normal soft tissue sample from the same or different individual, determining the expression of a gene in both samples and comparing the expression of the gene in both soft tissue samples, where a higher level of protein expression in the first soft tissue sample indicates the presence of soft tissue sarcoma. The method of the invention has cytostatic applications and may be useful for detecting soft tissue	

CC sarcoma, possibly via gene therapy or vaccine production. The nucleic acid sequences may be useful in diagnostic and screening applications. The current sequence is that of a human soft tissue sarcoma-upregulated DNA of the invention. The current sequence is not shown within the specification per se but was submitted in CD format by the inventor.

XX Sequence 259202 BP; 81699 A; 51421 C; 49221 G; 76861 T; 0 U; 0 Other;

Query Match 55.8%; Score 274.6; DB 12; Length 259202;

Best Local Similarity 79.0%; Pred. No. 1.6e-73;

Matches 394; Conservative 0; Mismatches 89; Indels 16; Gaps 5;

```

QY 1 ATGGGACCTGATTTTGTGCTGCTGCTGAGAGAGCTTGTGATGCGCTGAGT 60
DB 69873 ATGGGACCTGATTTTGTGCTGCTGCTGAGAGAGCTTGTGATGCGCTGAGT 69932
QY 61 AAAATACCCCT-----GCATTAATTCATTTTCAAGCTTGGAAAATAAGTCTGCC 114
DB 69933 AAAACACCCCTGATTAAGTCAAGTGTCCATTTTCAAACTTGGACATTAATCTGCTC 69992
QY 115 CACAGTTGTAACCTTTAAGGTTAAGCAGTACAAATGATGCTCTCAATGCTCT 174
DB 69993 ATAGTTGCTGAAA--TTAGGTTTAAACAGTATGAGTCAATGCTCTCAATGCTCT 70050
QY 175 GTGTTTAAGAAACACTTGAAGAGCTTGTATTAATAAATAATATATTCAGATGCTCC 234
DB 70051 GGGTTTAAAGAAACACTT--CAGGAGCTTGTTTAAAGATATATTCATATGCCGCTAC 70109
QY 235 ACCCAAGACTGATTCAGTAAAGCAGAGCTGGGGAGATGCCAGACTCTGCATTTTAA 294
DB 70110 CAAAATTTGATTTGTTGATCAGCTGGGGAGG-----GCCCAGAGCTCTGCATTTTAA 70163
QY 295 AAGCACTCAGAGATCTGTGAGAGCAATTAATCTGTAATAATCATGCGCCATCTCTAG 354
DB 70164 AAGCACTCAGAGATCTGTGAGAGCAATTAATCTGTAATAATCATGCGCCATCTCTAG 70223
QY 355 ATGAGAGAACTTTTGAAGAGGACCTTGAAGGCTTCCAGAGAAAGTGTCTGACAGCT 414
DB 70224 ATGAGAGAACTTTTGAAGAGGACCTTGAAGGCTTCCAGAGAAAGTGTCTTAAACAGCT 70283
QY 415 TAGG-CAAATCTCAAAAATGCGAATTTTCTTAATAACCAATTTTAAAGAGTGTCCA 473
DB 70284 TTGAGCAAAATTTTCAAGAGATGCGAATTTTGTCTTAATAACCAATTTTCAAGATTTCA 70343
QY 474 ACTCTCTCTGCGCTTCCA 492
DB 70344 AATCTCTCTGCGCTTCCA 70362

RESULT 2
ADM80835
ID ADM80835 standard; cDNA; 270 BP.
XX
AC ADM80835;
XX
DT 03-JUN-2004 (first entry)
DE Human CADECM-22 encoding cDNA SEQ ID NO:64.
XX
XX human; cell adhesion and extracellular matrix protein; CADECM;
KM neuroprotective; cytoskeletal; anorectic; immune disorder;
KW neurological disorder; developmental disorder;
KW connective tissue disorder; cell proliferative disorder; cancer; obesity;
KW tangler disease; gene; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 36..212
FT FT /*tag= a
FT FT /product= "CADECM-22"
PN MO2004015396-AA2.

```

```

XX
PD 19-FEB-2004.
XX
PF 12-AUG-2003; 2003MO-US025418.
XX
PR 13-AUG-2002; 2002US-0403781P.
PR 30-AUG-2002; 2002US-0407034P.
PR 13-SEP-2002; 2002US-0410566P.
PR 24-SEP-2002; 2002US-0413482P.
PR 25-SEP-2002; 2002US-0413890P.
PR 08-NOV-2002; 2002US-0424904P.
PR 13-NOV-2002; 2002US-0426222P.
XX
PA (INCY-) INCYTE CORP.
XX
PI Elliott VS, Khare R, Emerling BM, Kable AE, Tran UK, Jin P,
PI Becha SD, Margulis JP, Swarnakar A, Chawla NK, Ramkumar J,
PI Hafalia AJA, Lee SY, Jiang X, Jackson AA, Richardson TW, Blake JJ,
PI Wang JT, Chien D, Yang YG;
XX
DR WPI, 2004-191795/18.
DR P-PADB; ADM80793.
XX
PT New cell adhesion and extracellular matrix proteins, useful in
PT diagnosing, treating and preventing immune, neurological, developmental,
PT connective tissue and cell proliferative disorders including cancer.
XX
PS Claim 5; SEQ ID NO 64; 272bp; English.
XX
CC The present sequence encodes a human cell adhesion and extracellular
CC matrix protein designated CADECM. CADECM sequences has neuroprotective,
CC cytoskeletal and anorectic activities. The CADECM polypeptides and
CC polynucleotides are useful in diagnosing, treating and preventing immune,
CC neurological, developmental, connective tissue and cell proliferative
CC disorders including cancer, e.g. breast, prostate, ovarian, lung or colon
CC cancer, obesity and tangler disease.
XX
SQ Sequence 270 BP; 76 A; 65 C; 67 G; 62 T; 0 U; 0 Other;

Query Match 10.4%; Score 51.4; DB 12; Length 270;
Best Local Similarity 90.2%; Pred. No. 1.7e-05;
Matches 55; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 ATGGGACCTGATTTTGTGCTGCTGCTGAGAGAGCTTGTGATGCGCTGAGT 60
DB 36 ATGGGACCTGATTTTGTGCTGCTGCTGAGAGAGCTTGTGATGCGCTGAGT 95
QY 61 A 61
DB 96 A 96

RESULT 3
ADM80826
ID ADM80826 standard; cDNA; 549 BP.
XX
AC ADM80826;
XX
DT 03-JUN-2004 (first entry)
DE Human CADECM-13 encoding cDNA SEQ ID NO:55.
XX
XX human; cell adhesion and extracellular matrix protein; CADECM;
KM neuroprotective; cytoskeletal; anorectic; immune disorder;
KW neurological disorder; developmental disorder;
KW connective tissue disorder; cell proliferative disorder; cancer; obesity;
KW tangler disease; gene; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 36..491
FT FT /*tag= a
FT FT

```

/product= "CADECM-13"

FT XX WO2004015396-A2.  
FN XX  
PD XX 19-FEB-2004.  
PF XX 12-AUG-2003; 2003WO-US025418.  
PR XX 13-AUG-2002; 2002US-0403781P.  
PR 30-AUG-2002; 2002US-0407034P.  
PR 13-SEP-2002; 2002US-0410566P.  
PR 24-SEP-2002; 2002US-0413482P.  
PR 25-SEP-2002; 2002US-0413890P.  
PR 08-NOV-2002; 2002US-0424904P.  
PR 13-NOV-2002; 2002US-0426222P.  
XX  
XX (INCY-) INCYTE CORP.  
XX  
PI Elliott VS, Khare R, Emerling BM, Kable AE, Tran UK, Jin P,  
PI Becha SD, Margulis JP, Swarnakar A, Chawla NK, Ramkumar J,  
PI Hafalia AJA, Lee SY, Jiang X, Jackson AA, Richardson TW, Blake JJ,  
PI Wang JT, Chien D, Yang YG;  
DR WPI; 2004-191795/18.  
DR P-PSDB; ADM80784.  
XX  
XX New cell adhesion and extracellular matrix proteins, useful in  
PT diagnosing, treating and preventing immune, neurological, developmental,  
PT connective tissue and cell proliferative disorders including cancer.  
XX  
PS Claim 5; SEQ ID NO 55; 272pp; English.  
XX  
XX The present sequence encodes a human cell adhesion and extracellular  
CC matrix protein designated CADECM. CADECM sequences has neuroprotective,  
CC cytotactic and anorectic activities. The CADECM polypeptides and  
CC polynucleotides are useful in diagnosing, treating and preventing immune,  
CC neurological, developmental, connective tissue and cell proliferative  
CC disorders including cancer, e.g. breast, prostate, ovarian, lung or colon  
CC cancer, obesity and tanger disease.  
XX  
SQ Sequence 549 BP; 134 A; 203 C; 112 G; 100 T; 0 U; 0 Other;

Query Match 10.4%; Score 51.4; DB 12; Length 549;  
Best Local Similarity 90.2%; Pred. No. 2.3e-05;  
Matches 55; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 ATGGGACCTGATTTTGTGCTGCTCTGCGAGACGCTTTGCTATGCCGTGAGT 60  
DB |||||  
36 ATGGGACCTGATTTTGTGCTGCTCTGCGAGACGCTTTGCTATGCCGTGAGT 95

QY 61 A 61  
DB 96 A 96

RESULT 4  
ADM80827  
ID ADM80827 standard; cDNA; 623 BP.  
XX  
AC ADM80827;  
XX  
DT 03-JUN-2004 (first entry)  
XX  
DE Human CADECM-14 encoding cDNA SEQ ID NO:56.  
XX  
XX human; cell adhesion and extracellular matrix protein; CADECM;  
XX neuroprotective; cytotactic; anorectic; immune disorder;  
XX neurological disorder; developmental disorder;  
XX connective tissue disorder; cell proliferative disorder; cancer; obesity;  
XX tanger disease; gene; ss.  
OS Homo sapiens.  
XX

Key Location/Qualifiers  
FT CDS 36..563  
FT /\*tag= a  
FT /product= "CADECM-14"

FT XX WO2004015396-A2.  
FN XX  
PD XX 19-FEB-2004.  
PF XX 12-AUG-2003; 2003WO-US025418.  
PR XX 13-AUG-2002; 2002US-0403781P.  
PR 30-AUG-2002; 2002US-0407034P.  
PR 13-SEP-2002; 2002US-0410566P.  
PR 24-SEP-2002; 2002US-0413482P.  
PR 25-SEP-2002; 2002US-0413890P.  
PR 08-NOV-2002; 2002US-0424904P.  
PR 13-NOV-2002; 2002US-0426222P.  
XX  
XX (INCY-) INCYTE CORP.  
XX  
PI Elliott VS, Khare R, Emerling BM, Kable AE, Tran UK, Jin P,  
PI Becha SD, Margulis JP, Swarnakar A, Chawla NK, Ramkumar J,  
PI Hafalia AJA, Lee SY, Jiang X, Jackson AA, Richardson TW, Blake JJ,  
PI Wang JT, Chien D, Yang YG;  
DR WPI; 2004-191795/18.  
DR P-PSDB; ADM80785.  
XX  
XX New cell adhesion and extracellular matrix proteins, useful in  
PT diagnosing, treating and preventing immune, neurological, developmental,  
PT connective tissue and cell proliferative disorders including cancer.  
XX  
PS Claim 5; SEQ ID NO 56; 272pp; English.  
XX  
XX The present sequence encodes a human cell adhesion and extracellular  
CC matrix protein designated CADECM. CADECM sequences has neuroprotective,  
CC cytotactic and anorectic activities. The CADECM polypeptides and  
CC polynucleotides are useful in diagnosing, treating and preventing immune,  
CC neurological, developmental, connective tissue and cell proliferative  
CC disorders including cancer, e.g. breast, prostate, ovarian, lung or colon  
CC cancer, obesity and Tanger disease.  
XX  
SQ Sequence 623 BP; 150 A; 228 C; 128 G; 117 T; 0 U; 0 Other;

Query Match 10.4%; Score 51.4; DB 12; Length 623;  
Best Local Similarity 90.2%; Pred. No. 2.4e-05;  
Matches 55; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 ATGGGACCTGATTTTGTGCTGCTCTGCGAGACGCTTTGCTATGCCGTGAGT 60  
DB |||||  
36 ATGGGACCTGATTTTGTGCTGCTCTGCGAGACGCTTTGCTATGCCGTGAGT 95

QY 61 A 61  
DB 96 A 96

RESULT 5  
AAZ50832  
ID AAZ50832 standard; DNA; 476 BP.  
XX  
AC AAZ50832;  
XX  
DT 31-MAY-2000 (first entry)  
XX  
DE Rat amelogenin gene (A4).  
XX  
XX Amelogenin; splice variant; rat; (A4); chondrogenesis; osteogenesis;  
XX chondrogenic inducing molecule; CIM; cartilage growth; osteopathic;  
XX extracellular matrix protein; tooth enamel; enamel mineralisation;  
XX ameloblast; bone regeneration; composite cell construct; ds.  
XX

```
OS Rattus sp.
XX Key Location/Qualifiers
XX exon 1..36
XX /tag= a
XX /partial
XX /number= 1
XX exon 37..101
XX /tag= b
XX /number= 2
XX CDS 48..317
XX /tag= c
XX /product= "Rat amelogenin protein"
XX /transl_except= (pos:69..71, aa:GLY)
XX sig_peptide 48..95
XX /tag= d
XX mat_peptide 96..314
XX /tag= e
XX /label= Mature_rat_amelogenin_protein
XX exon 102..149
XX /tag= f
XX /number= 3
XX exon 150..191
XX /tag= g
XX /number= 4
XX exon 192..236
XX /tag= h
XX /number= 5
XX exon 237..311
XX /tag= i
XX /number= 6d
XX /note= "Comprises of gene segments 6a, b, c and d"
XX exon 312..317
XX /tag= j
XX /number= 7
XX /note= "Includes the stop codon"
XX WO200006734-A1.
XX 10-FEB-2000.
XX 29-JUL-1999; 99WO-US017342.
XX 29-JUL-1998; 98US-0094489P.
XX (NOUN ) UNIV NORTHWESTERN.
XX Veis A, Nebgen DR;
XX PFI; 2000-205464/18.
XX P-PSDB; AAY45074.
XX Novel amelogenin polypeptides and polynucleotides, useful for enhancing
XX bone generation in mammals and synthesizing bone matrix or articular
XX surfaces at implant sites.
XX Example 2; Fig 11B; 79P; English.
XX The present DNA sequence is the full-length rat amelogenin gene (A4),
XX comprising exons 1-7, including the exon segment 6d. It is derived from
XX the rat incisor odontoblast-pulp cDNA library. The splice variants of
XX this gene functions as an osteogenic or chondrogenic inducing molecule
XX (CIM), which is useful for enhancing bone or cartilage growth. It has
XX osteopathic activity. Amelogenin belongs to the family of extracellular
XX matrix proteins, in developing tooth enamel, that are produced by the
XX ameloblasts and plays a role in enamel mineralisation. Chondrogenic or
XX osteogenic inducing amelogenin molecules are useful to induce
XX differentiation of cells to the osteogenic and chondrogenic phenotypes
XX and can be used in a composite cell construct for bone and cartilage
XX regeneration. The polynucleotides can be employed to produce the
XX polypeptides by recombinant techniques
XX Sequence 476 BP; 155 A; 106 C; 98 G; 117 T; 0 U; 0 Other;
```

```
Query Match 10.4%; Score 51.2; DB 3; Length 476;
Best Local Similarity 94.6%; Pred. No. 2.5e-05;
Matches 53; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 ATGGGACCGGATTTGTCGCTCCCTCGGGAGCAGCCCTTGCTATGCCCT 56
DB 48 ATGGGACCTGAGTCTTGTCCTGCTCGGGAGCAGCTTGTATGCCCT 103
RESULT 6
ADB59026
ID ADB59026 standard; DNA; 556 BP.
XX ADB59026;
XX 04-DEC-2003 (first entry)
XX Toxicity-related gene, SEQ ID 4052.
XX Toxic; toxin; gene expression profile; hepatotoxicity; liver;
XX drug screening; toxicity assay; ds.
XX Unidentified.
XX MO2003064624-A2.
XX 07-AUG-2003.
XX 31-JAN-2003; 2003WO-US003194.
XX 31-JAN-2002; 2002US-00060087.
XX 15-MAR-2002; 2002US-0364045P.
XX 15-MAR-2002; 2002US-0364055P.
XX 30-DEC-2002; 2002US-0436643P.
XX (GENE-) GENE LOGIC INC.
XX Mendrick D, Porter M, Johnson K, Higge B, Castle A, Elashoff M;
XX WPI; 2003-689530/65.
XX Predicting a toxic effect of a compound, useful in identifying toxicity
XX markers in liver tissues or cells for drug screening and toxicity assays,
XX comprises preparing gene expression profile of tissue or cells exposed to
XX the compound.
XX Claim 1; SEQ ID NO 4052; 1156bp; English.
XX The present invention relates to a method for predicting a toxic effect
XX of a compound. The method comprises preparing a gene expression profile
XX of a tissue or cell sample exposed to the compound, and comparing the
XX gene expression profile to a database comprising SEQ ID 1-4925, where
XX differential expression of the gene indicates at least one toxic effect.
XX The method is useful for predicting at least one toxic effect of a
XX compound, predicting hepatotoxicity or the progression of a toxic effect
XX of a compound, identifying an agent that modulates the onset or
XX progression of a toxic response, predicting the cellular pathways that a
XX compound modulates in a cell, and identifying an agent that modulates at
XX least one activity of a protein. The method and compositions of the
XX present invention using a database of genes having liver toxin-induced
XX differential expression, are useful in identifying toxicity markers in
XX liver tissues or cells for drug screening and toxicity assays. Note: The
XX sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 556 BP; 121 A; 183 C; 114 G; 119 T; 0 U; 19 Other;
Query Match 10.2%; Score 50.2; DB 10; Length 556;
Best Local Similarity 92.9%; Pred. No. 5.4e-05;
Matches 52; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

OY 1 ATGGGACCTGATTTGCTGCTGCTGAGACGCTTGTATGCCCGT 56  
DB 22 ATGGGACCTGATCTGTTGCTGCTGCTGAGACGCTTGTATGCCCGT 77

RESULT 7  
ID ADB53782 standard; DNA; 556 BP.  
AC ADB53782;  
XX  
DT 04-DEC-2003 (first entry)  
XX  
DE Primary rat hepatocyte toxicity modelling related gene SEQ ID NO:4324.  
XX  
KM toxic effect; gene expression profile; hepatotoxicity; diagnostic marker;  
KM toxicity marker; toxicity progression; drug screening;  
KM primary rat hepatocyte toxicity modelling; gene; ds.  
XX  
OS Rattus norvegicus.  
XX  
PN WO2003065993-A2.  
XX  
PD 14-AUG-2003.  
XX  
PF 04-FEB-2003; 2003NO-US003482.  
XX  
PR 04-FEB-2002; 2002US-0353171P.  
XX  
PR 13-MAR-2002; 2002US-0363534P.  
XX  
PR 08-APR-2002; 2002US-0370248P.  
XX  
PR 10-APR-2002; 2002US-0371134P.  
XX  
PR 10-APR-2002; 2002US-0371135P.  
XX  
PR 10-APR-2002; 2002US-0371150P.  
XX  
PR 11-APR-2002; 2002US-0371413P.  
XX  
PR 19-APR-2002; 2002US-0373601P.  
XX  
PR 22-APR-2002; 2002US-0374139P.  
XX  
PR 08-MAY-2002; 2002US-0378370P.  
XX  
PR 09-MAY-2002; 2002US-0378652P.  
XX  
PR 09-MAY-2002; 2002US-0378653P.  
XX  
PR 09-MAY-2002; 2002US-0378655P.  
XX  
PR 09-JUL-2002; 2002US-0394230P.  
XX  
PR 09-JUL-2002; 2002US-0394253P.  
XX  
PR 04-SRP-2002; 2002US-0407688P.  
XX  
PR 28-JAN-2003; 2003US-0442900P.  
XX  
PA (GENE-) GENE LOGIC INC.  
XX  
PI Mendrick D, Porter M, Johnson K, Higgs B, Caastle A, Orr M;  
PI Elashoff M;  
XX  
DR WPI; 2003-731472/69.  
XX  
PT Determining if a compound induces a toxic effect on a tissue or cell, for  
PT identifying hepatotoxic compounds, comprises comparing a gene expression  
PT profile of a tissue or cell sample to a database of Tox mean and non-Tox  
PT mean values.  
XX  
PS Claim 44; SEQ ID NO 4324; 874bp; English.  
XX  
CC The present invention describes a method for determining whether a  
CC compound induces a toxic effect on a tissue or cell. The method comprises  
CC preparing a gene expression profile of a tissue or cell sample exposed to  
CC the compound, and comparing the gene expression profile to a database  
CC comprising data or information on the Tox mean and non-Tox mean value.  
CC The method is useful for predicting or identifying at least one toxic  
CC effect, particularly hepatotoxicity, of a test or unknown compound. The  
CC genes listed in the specification are useful as diagnostic or toxicity  
CC markers for the prediction or identification of the physiological state  
CC of tissue or cell sample that has been exposed to a compound, or to  
CC identify or predict the toxic effects of a compound or an agent. These  
CC can also be used as markers for monitoring toxicity progression or for  
CC drug screening. The present sequence represents a primary rat hepatocyte

CC toxicity modelling related gene sequence from the present invention.  
XX  
SQ Sequence 556 BP; 121 A; 183 C; 114 G; 119 T; 0 U; 19 Other;  
XX

Query Match 10.2%; Score 50.2; DB 10;  
Best Local Similarity 92.9%; Pred. No. 5,4e-05;  
Matches 52; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 ATGGGACCTGATTTGCTGCTGCTGAGACGCTTGTATGCCCGT 56  
DB 22 ATGGGACCTGATCTGTTGCTGCTGCTGAGACGCTTGTATGCCCGT 77

RESULT 8  
ID AAL41111 standard; DNA; 722 BP.  
XX  
AC AAL41111;  
XX  
DT 16-OCT-2002 (first entry)  
XX  
DE gAML related Y-chromosome DNA sequence.  
XX  
KM Goat embryo sexual identification technique; goat amelogenin gene; gAML;  
KM sex-specific; gene; ds; Y-chromosome.  
XX  
OS Capra hircus.  
XX  
XX  
XX Key Location/Qualifiers  
XX FT 35..658  
XX FT CDS /\*tag= a  
XX FT /product= "Y-chromosome protein"  
XX  
XX TW454013-A.  
XX  
PD 11-SEP-2001.  
XX  
PF 10-NOV-1999; 99TW-00119616.  
XX  
PR 10-NOV-1999; 99TW-00119616.  
XX  
XX 10-NOV-1999; 99TW-00119616.  
XX  
XX (CHEN/) CHEN C.  
XX (JANG/) JANG J.  
XX (WENG/) WENG T.  
XX (JENG/) JENG D.  
XX  
XX Chen C, Jang J, Weng T, Jeng D;  
XX  
XX WPI; 2002-442016/47.  
XX  
DR P-PSDB; AAO22534.  
XX  
PT Sex-specific sequence of goat amelogenin gene, useful for embryo sexual  
PT identification, comprises high sensitivity even using single white blood  
PT cell or cleavage c.  
XX  
PS Disclosure; Page 28; 35bp; Chinese.  
XX  
XX The invention relates to a goat embryo sexual identification technique  
XX with high efficiency, sensitivity and repeatability. This technique  
XX involves separately cloning and sequencing the coding regions and the  
XX introns of the goat amelogenin gene (gAML) on the goat chromosomes. The  
XX results indicate that there are sex-specific sequences in the fifth  
XX intron of the gene. The major characteristics according to the present  
XX invention include high sensitivity, applicable in sex identification even  
XX only using a single white blood cell or a single cleavage cell of  
XX blastula; high diagnostic efficiency, capable of identifying hundreds of  
XX goat embryo in 3 hours; simple operation procedures without complicated  
XX steps of DNA extraction and need no additional control group intron; and  
XX can be applied on different species of goats. This polynucleotide  
XX sequence represents a gAML related Y-chromosome DNA sequence of the  
XX  
SQ Sequence 722 BP; 173 A; 264 C; 140 G; 145 T; 0 U; 0 Other;

Query Match 10.1%; Score 49.8; DB 6; Length 722;  
Best Local Similarity 96.2%; Pred. No. 8e-05; Mismatches 2; Indels 0; Gaps 0;  
Matches 51; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ATGGGAGACCTGATTTGTTGCTGCTGCTGAGAGAGCTTGTATGCC 53  
Db 35 ATGGGAGACCTGATTTGTTGCTGCTGCTGAGAGAGCTTGTATGCC 87

## RESULT 9

AL41110  
ID AL41110 standard; DNA; 752 BP.

AC AAL41110;

DT 16-OCT-2002 (first entry)

DE gAML related X-chromosome DNA sequence.

KM Goat embryo sexual identification technique; goat amelogenin gene; gAML;

OS Capra hircus.

Key Location/Qualifiers

FT CDS 35..658

FT /tag= a

FT /product= "X-chromosome protein"

PN TW454013-A.

PD 11-SEP-2001.

PE 10-NOV-1999; 99TW-00119616.

PR 10-NOV-1999; 99TW-00119616.

PA (CHEN/) CHEN C.

PA (JANG/) JANG J.

PA (WENG/) WENG T.

PA (JENG/) JENG D.

PI Chen C, Jang J, Weng T, Jeng D;

DR WPI: 2002-442016/47.

DR P-PSDB; AAO22534.

PT Sex-specific sequence of goat amelogenin gene, useful for embryo sexual

PT identification, comprises high sensitivity even using single white blood

PT cell or cleavage c.

PS Disclosure; Page 28; 35pp; Chinese.

XX The invention relates to a goat embryo sexual identification technique

XX with high efficiency, sensitivity and repeatability. This technique

XX involves separately cloning and sequencing the coding regions and the

XX introns of the goat amelogenin gene (gAML) on the goat chromosomes. The

XX results indicate that there are sex-specific sequences in the fifth

XX intron of the gene. The major characteristics according to the present

XX invention include high sensitivity, applicable in sex identification even

XX only using a single white blood cell or a single cleavage cell of

XX blastula; high diagnostic efficiency, capable of identifying hundreds of

XX goat embryo in 3 hours; simple operation procedures without complicated

XX steps of DNA extraction and need no additional control group intron; and

XX can be applied on different species of goats. This polynucleotide

XX sequence represents a gAML related X-chromosome DNA sequence of the

XX invention

XX Sequence 752 BP; 184 A; 271 C; 152 G; 145 T; 0 U; 0 Other;

XX Query Match 10.1%; Score 49.8; DB 6; Length 752;

XX Best Local Similarity 96.2%; Pred. No. 8.2e-05;

Matches 51; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ATGGGAGACCTGATTTGTTGCTGCTGCTGAGAGAGCTTGTATGCC 53  
Db 35 ATGGGAGACCTGATTTGTTGCTGCTGCTGAGAGAGCTTGTATGCC 87

## RESULT 10

ADM80836  
ID ADM80836 standard; cDNA; 318 BP.

AC ADM80836;

DT 03-JUN-2004 (first entry)

DE Human CADECM-23 encoding cDNA SEQ ID NO:65.

KM human; cell adhesion and extracellular matrix protein; CADECM;

KM neuroprotective; cytoskeletal; anorectic; immune disorder;

KM neurological disorder; developmental disorder;

KM connective tissue disorder; cell proliferative disorder; cancer; obesity;

OS Homo sapiens.

Key Location/Qualifiers

FT CDS 36..260

FT /tag= a

FT /product= "CADECM-23"

PN WO2004015396-A2.

PD 19-FEB-2004.

PE 12-ANG-2003; 2003WO-US025418.

PR 13-AUG-2002; 2002US-0403781P.

PR 30-SEP-2002; 2002US-0407034P.

PR 13-SEP-2002; 2002US-0410566P.

PR 24-SEP-2002; 2002US-0413482P.

PR 25-SEP-2002; 2002US-0413890P.

PR 08-NOV-2002; 2002US-0424904P.

PR 13-NOV-2002; 2002US-0426222P.

PA (INCY-) INCYTE CORP.

PI Elliott VS, Khare R, Emerling BM, Kadle AE, Tran UK, Jin P,

PI Becha SD, Margulis JP, Swarnakar A, Chawla NK, Ramkumar J,

PI Hafalia AJA, Lee SY, Jhang X, Jackson AA, Richardson TW, Blake JJ,

PI Wang JT, Chien D, Yang YG;

DR WPI: 2004-191795/18.

DR P-PSDB; ADM80794.

PT New cell adhesion and extracellular matrix proteins, useful in

PT diagnosing, treating and preventing immune, neurological, developmental,

PT connective tissue and cell proliferative disorders including cancer.

PS Claim 5; SEQ ID NO 65; 272pp; English.

XX The present sequence encodes a human cell adhesion and extracellular

XX matrix protein designated CADECM. CADECM sequences has neuroprotective,

XX cytoskeletal and anorectic activities. The CADECM polypeptides and

XX polynucleotides are useful in diagnosing, treating and preventing immune,

XX neurological, developmental, connective tissue and cell proliferative

XX disorders including cancer, e.g. breast, prostate, ovarian, lung or colon

XX cancer, obesity and tangler disease.

XX Sequence 318 BP; 87 A; 81 C; 75 G; 75 T; 0 U; 0 Other;

XX Query Match 9.8%; Score 48.2; DB 12; Length 318;

XX Best Local Similarity 94.3%; Pred. No. 0.00018;

XX Matches 50; Conservative 0; Mismatches 3; Indels 0; Gaps 0;



OY 1 ATGGGACCTGATTTGCTGCTGCTGAGACGCTTGTATGCC 53  
DB 36 ATGGGACCTGATTTATTTGCTGCTGCTGAGACGCTTGTGCAATGCC 88

RESULT 11  
AAZ07020  
ID AAZ07020 standard; DNA; 750 BP.

AC AAZ07020;

DT 15-NOV-1999 (first entry)

DE Amelogenin X nucleotide sequence.

XX Amelogenin X; Amelogenin Y; genotype; 5' nuclease amplification;  
KM multiple allelic site; apolipoprotein E; apoE; coronary artery disease;  
XX Alzheimer's disease; db.

OS Unidentified.

PN WO940226-A2.

PD 12-AUG-1999.

PF 08-JAN-1999; 99WO-US000499.

PR 04-FEB-1998; 98US-00018595.

PA (PEKE ) PERKIN-ELMER CORP.

PI Liyak KJ, Goodsaalid F;

DR WPI; 1999-539985/45.

PT 5' nuclease amplification assay using fluorescence-quencher probes for  
determination of a genotype at multiple allelic sites.

PS Disclosure; Fig 10; 95pp; English.

XX The present invention describes first and second sets of fluorescer-  
quencher probes used simultaneously in a 5' nuclease assay to identify  
CC which members of a first or second set of substantially homologous  
CC sequences are present in a DNA sample. The method can be used to genotype  
CC a sample of genomic DNA at two or more different allelic sites.  
CC Generating a fluorescence spectrum and signature for each genotype, which  
CC uniquely reflects the assay's inherent inefficiency for that genotype  
CC given the particular conditions, probes and primers used, the genotype of  
CC unknown sequences can be determined. The assay was shown to be useful for  
CC determining apoe genotypes. The assay can be used as a diagnostic tool  
CC for assessing the risk for coronary artery disease and/or late-onset  
CC Alzheimer's disease. Using the 5' nuclease assay of the invention it is  
CC possible to determine a genotype at two or more allelic sites in a single  
CC reaction. This approach is much faster than previous approaches to  
CC genotyping genes having two or more allelic sites, such as the  
CC apolipoprotein E gene. A key advantage of the method for determining the  
CC genotype of a sample of DNA at multiple allelic sites is that it does not  
CC rely on 5' nuclease assay working with 100% efficiency to distinguish  
CC between substantially homologous sequences such as alleles. The present  
CC sequence represents the nucleotide sequence for amelogenin X, which is  
CC used in the exemplification of the present invention

XX Sequence 750 BP; 191 A; 260 C; 146 G; 153 T; 0 U; 0 Other;

Query Match 9.8%; Score 48.2; DB 2; Length 750;

Best Local Similarity 94.3%; Pred. No. 0.00026;  
Matches 50; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 ATGGGACCTGATTTGCTGCTGCTGAGACGCTTGTATGCC 53  
DB 69 ATGGGACCTGATTTATTTGCTGCTGCTGAGACGCTTGTGCAATGCC 121

RESULT 12  
AAZ07018  
ID AAZ07018 standard; DNA; 793 BP.

AC AAZ07018;

DT 15-NOV-1999 (first entry)

DE Amelogenin X nucleotide sequence.

XX Amelogenin X; Amelogenin Y; genotype; 5' nuclease amplification;  
KM multiple allelic site; apolipoprotein E; apoE; coronary artery disease;  
XX Alzheimer's disease; db.

OS Unidentified.

PN WO940226-A2.

PD 12-AUG-1999.

PF 08-JAN-1999; 99WO-US000499.

PR 04-FEB-1998; 98US-00018595.

PA (PEKE ) PERKIN-ELMER CORP.

PI Liyak KJ, Goodsaalid F;

DR WPI; 1999-539985/45.

PT 5' nuclease amplification assay using fluorescence-quencher probes for  
determination of a genotype at multiple allelic sites.

PS Disclosure; Fig 8A; 95pp; English.

XX The present invention describes first and second sets of fluorescer-  
quencher probes used simultaneously in a 5' nuclease assay to identify  
CC which members of a first or second set of substantially homologous  
CC sequences are present in a DNA sample. The method can be used to genotype  
CC a sample of genomic DNA at two or more different allelic sites.  
CC Generating a fluorescence spectrum and signature for each genotype, which  
CC uniquely reflects the assay's inherent inefficiency for that genotype  
CC given the particular conditions, probes and primers used, the genotype of  
CC unknown sequences can be determined. The assay was shown to be useful for  
CC determining apoe genotypes. The assay can be used as a diagnostic tool  
CC for assessing the risk for coronary artery disease and/or late-onset  
CC Alzheimer's disease. Using the 5' nuclease assay of the invention it is  
CC possible to determine a genotype at two or more allelic sites in a single  
CC reaction. This approach is much faster than previous approaches to  
CC genotyping genes having two or more allelic sites, such as the  
CC apolipoprotein E gene. A key advantage of the method for determining the  
CC genotype of a sample of DNA at multiple allelic sites is that it does not  
CC rely on 5' nuclease assay working with 100% efficiency to distinguish  
CC between substantially homologous sequences such as alleles. The present  
CC sequence represents the nucleotide sequence for amelogenin X, which is  
CC used in the exemplification of the present invention

XX Sequence 793 BP; 215 A; 266 C; 150 G; 162 T; 0 U; 0 Other;

Query Match 9.8%; Score 48.2; DB 2; Length 793;

Best Local Similarity 94.3%; Pred. No. 0.00026;  
Matches 50; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 ATGGGACCTGATTTGCTGCTGCTGAGACGCTTGTATGCC 53  
DB 69 ATGGGACCTGATTTATTTGCTGCTGCTGAGACGCTTGTGCAATGCC 121

RESULT 13  
AAZ07019  
ID AAZ07019 standard; DNA; 802 BP.

AC AA207019;  
XX  
DT 15-NOV-1999 (first entry)  
XX  
DE Amelogenin Y nucleotide sequence.  
XX  
KM Amelogenin X; Amelogenin Y; genotype: 5' nuclease amplification;  
KM multiple allelic site; apolipoprotein E, apoE; coronary artery disease;  
KM Alzheimer's disease; ss.  
XX  
OS unidentified.  
XX  
PN WO9940226-A2.  
XX  
PD 12-AUG-1999.  
XX  
PF 08-JAN-1999; 99WO-US000499.  
XX  
PR 04-FEB-1998; 98US-00018595.  
XX  
PA (PEKE) PERKIN-ELMER CORP.  
XX  
PI Livak KJ, Goodaaid F;  
XX  
DR WPI; 1999-539985/45.  
XX  
PT 5' nuclease amplification assay using fluorescence-quencher probes for  
PT determination of a genotype at multiple allelic sites.  
XX  
PS Disclosure; Fig 8B; 95pp; English.  
XX  
The present invention describes first and second sets of fluorescent-  
quencher probes used simultaneously in a 5' nuclease assay to identify  
CC which members of a first or second set of substantially homologous  
CC sequences are present in a DNA sample. The method can be used to genotype  
CC a sample of genomic DNA at two or more different allelic sites.  
CC Generating a fluorescence spectrum and signature for each genotype, which  
CC uniquely reflects the assay's inherent inefficiency for that genotype  
CC given the particular conditions, probes and primers used, the genotype of  
CC unknown sequences can be determined. The assay was shown to be useful for  
CC determining apolipoprotein E genotype. The assay can be used as a diagnostic tool  
CC for assessing the risk for coronary artery disease and/or late-onset  
CC Alzheimer's disease. Using the 5' nuclease assay of the invention it is  
CC possible to determine a genotype at two or more allelic sites in a single  
CC reaction. This approach is much faster than previous approaches to  
CC genotyping genes having two or more allelic sites, such as the  
CC apolipoprotein E gene. A key advantage of the method for determining the  
CC genotype of a sample of DNA at multiple allelic sites is that it does not  
CC rely on 5' nuclease assay working with 100% efficiency to distinguish  
CC between substantially homologous sequences such as alleles. The present  
CC sequence represents the nucleotide sequence for amelogenin Y, which is  
CC used in the exemplification of the present invention  
XX  
SQ Sequence 802 BP; 216 A; 258 C; 155 G; 173 T; 0 U; 0 Other;  
Query Match 9.5%; Score 46.6; DB 2; Length 802;  
Best Local Similarity 92.5%; Pred. No. 0.00083;  
Matches 49; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 ATGGGACCTGATTTTGTTCCTGCTCTCGGAGACACCTTTGCTATGCC 53  
DB 69 ATGGGACCTGATTTTGTTCCTGCTCTCGGAGACACCTTTGCTATGCC 121  
RESULT 14  
ADQ22977  
ID ADQ22977 standard; DNA; 852 BP.  
XX  
AC ADQ22977;  
XX  
DT 26-AUG-2004 (first entry)  
XX  
DE Human soft tissue sarcoma-upregulated DNA - SEQ ID 5797.

XX  
KM soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;  
KM da.  
XX  
OS Homo sapiens.  
XX  
PN WO2004048938-A2.  
XX  
PD 10-JUN-2004.  
XX  
PF 26-NOV-2003; 2003WO-US038193.  
XX  
PR 26-NOV-2002; 2002US-0429739P.  
XX  
PA (PROT-) PROTEIN DESIGN LABS INC.  
XX  
PI Aziz N, Ginsburg WM, Zlotnick A;  
XX  
DR WPI; 2004-441208/41.  
XX  
PT Early detection of soft tissue sarcoma comprises determining expression  
PT of a gene in a first soft tissue sample and a normal soft tissue sample  
PT and comparing the gene expression, also useful in treating soft tissue  
PT sarcoma.  
XX  
PS Example 2; SEQ ID NO 5797; 210pp; English.  
XX  
XX The invention relates to a novel method for detecting soft tissue sarcoma  
CC which comprises obtaining a first soft tissue sample from an individual  
CC and a normal soft tissue sample from the same or different individual,  
CC determining the expression of a gene in both samples and comparing the  
CC expression of the gene in both soft tissue samples, where a higher level  
CC of protein expression in the first soft tissue sample indicates the  
CC presence of soft tissue sarcoma. The method of the invention has  
CC cytostatic applications and may be useful for detecting soft tissue  
CC sarcoma, possibly via gene therapy or vaccine production. The nucleic  
CC acid sequences may be useful in diagnostic and screening applications.  
CC The current sequence is that of a human soft tissue sarcoma-upregulated  
CC DNA of the invention. The current sequence is not shown within the  
CC specification per se but was submitted in CD format by the inventor.  
XX  
SQ Sequence 852 BP; 226 A; 267 C; 169 G; 190 T; 0 U; 0 Other;  
Query Match 9.5%; Score 46.6; DB 12; Length 852;  
Best Local Similarity 92.5%; Pred. No. 0.00085;  
Matches 49; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 ATGGGACCTGATTTTGTTCCTGCTCTCGGAGACACCTTTGCTATGCC 53  
DB 69 ATGGGACCTGATTTTGTTCCTGCTCTCGGAGACACCTTTGCTATGCC 121  
RESULT 15  
ABN28389  
ID ABN28389 standard; DNA; 65 BP.  
XX  
AC ABN28389;  
XX  
DT 15-JUL-2002 (first entry)  
XX  
DE Rat spliced transcript detection oligonucleotide SEQ ID NO:1137.  
XX  
KM Human; mouse; rat; splice transcript; detection; RNA transcript;  
KM splice variant; transcriptome; oligonucleotide library; ss.  
XX  
OS Rattus norvegicus.  
XX  
PN WO200210449-A2.  
XX  
PD 07-FEB-2002.  
XX  
DT 20-JUL-2001; 2001WO-IB001903.  
XX

Wed Feb 9 09:28:09 2005

```

PR 28-JUL-2000; 2000US-0221607P.
PR 02-MAY-2001; 2001US-0267724P.
XX
PA (COMP-) COMPUGEN INC.
XX
PI Shoshan A, Messerman A, Mintz E, Mintz L, Faigler S;
XX
DR WPI; 2002-257383/30.
XX
PT New oligonucleotide libraries comprising oligonucleotides which
PT selectively hybridize to mRNAs transcribed from a transcription unit of a
PT genome, useful for detecting tissue-, pathology-, and developmental-
PT specific genes.
XX
PS Example 1; SEQ ID NO 1137; 47bp; English.
XX
CC The present invention describes oligonucleotide libraries for detecting
CC messenger RNAs that populate a (sub-)transcriptome, where the (sub-)
CC transcriptome comprises messenger RNAs transcribed from multiple
CC transcription units that populate a genome. The library comprises several
CC oligonucleotides, each capable of hybridizing selectively to a set of
CC messenger RNAs transcribed from a given transcription unit of the genome,
CC which encodes one or more messenger RNA splice variants. The
CC oligonucleotide libraries are useful for detecting mRNAs from a
CC biological sample, in expression profiling studies, in qualitatively or
CC quantitatively characterizing the corresponding transcriptome, and in
CC detecting RNA transcripts and splice variants of human or animal
CC transcriptomes. The libraries may also be used as specialised mini
CC libraries to detect transcripts of a sub-transcriptome under a particular
CC biological or pathological state, and so allowing the detection of tissue
CC - and pathology-specific genes such as those genes only expressed in
CC specific tissue under a specific pathological condition; to detect
CC developmental specific genes; and to detect RNA transcripts and splice
CC variants of a transcriptome of a patient suffering from a particular
CC disorder. ABN27253 to ABN59589 represent oligonucleotide sequences from
CC rats, humans and mice, which are used in the exemplification of the
CC present invention. N.B. The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 65 BP; 8 A; 23 C; 14 G; 20 T; 0 U; 0 Other;

```

Query Match 9.24; Score 45.2; DB 6; Length 65;  
Best Local Similarity 94.0%; Pred. No. 0.00079;  
Matches 47; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

QY 7 ACCTGGATTGTTGCTGCTCTCGAGCAGACCTTGTCTATGCCCGT 56
   |||||
DB 1 ACCTGGATTGTTGCTGCTCTCGAGCAGACCTTGTCTATGCCCGT 50

```

Search completed: February 8, 2005, 13:57:21  
Job time : 656.659 sec

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 8, 2005, 13:03:55 ; Search time 132.88 Seconds  
(without alignments)  
2631.757 Million cell updates/sec

Title: US-10-754-437-22

Perfect score: 492  
Sequence: 1 atggggagcctggatctgtc.....aactctctctgctccca 492

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 824507 seqs, 355394441 residues

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Issued Patents NA: \*  
1: /cgn2\_6/ptodate/1/ina/5A COMB.seq: \*  
2: /cgn2\_6/ptodate/1/ina/5B COMB.seq: \*  
3: /cgn2\_6/ptodate/1/ina/6A COMB.seq: \*  
4: /cgn2\_6/ptodate/1/ina/6B COMB.seq: \*  
5: /cgn2\_6/ptodate/1/ina/PTCUS COMB.seq: \*  
6: /cgn2\_6/ptodate/1/ina/backfile1.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	51.2	10.4	476	4 US-09-744-128-3	Sequence 3, Appl1
2	48.2	9.8	500	2 US-09-018-595B-3	Sequence 3, Appl1
3	48.2	9.8	500	3 US-09-324-709A-3	Sequence 3, Appl1
4	48.2	9.8	793	2 US-09-018-595B-1	Sequence 1, Appl1
5	48.2	9.8	793	3 US-09-324-709A-1	Sequence 1, Appl1
6	46.6	9.5	802	2 US-09-018-595B-2	Sequence 2, Appl1
7	46.6	9.5	802	3 US-09-324-709A-2	Sequence 2, Appl1
8	41.6	8.5	1141	4 US-09-806-708B-22	Sequence 22, Appl1
9	35.8	7.3	374	4 US-08-956-171E-4303	Sequence 4303, Ap
10	35.8	7.3	374	4 US-08-781-986A-4303	Sequence 4303, Ap
11	35.8	7.3	742	4 US-08-956-171E-3662	Sequence 3662, Ap
12	35.8	7.3	742	4 US-08-781-986A-3662	Sequence 3662, Ap
13	35.8	7.3	10758	4 US-08-956-171E-221	Sequence 221, App
14	35.8	7.3	10758	4 US-08-781-986A-221	Sequence 221, App
15	35	7.1	370	4 US-09-621-976-399	Sequence 399, App
16	34.2	6.9	832	4 US-09-621-976-2813	Sequence 2813, Ap
17	33.8	6.9	832	4 US-09-621-976-2813	Sequence 2813, Ap
18	33.8	6.9	1317	4 US-09-248-796A-6415	Sequence 6415, Ap
19	33.6	6.8	474	4 US-09-621-976-18033	Sequence 18033, A
20	33.4	6.8	399	4 US-09-621-976-8976	Sequence 8976, Ap
21	33.4	6.8	16063	3 US-09-801-052-3	Sequence 3, Appl1
22	33.4	6.8	414	4 US-10-020-121-3	Sequence 3, Appl1
23	33.2	6.7	2896	4 US-09-710-279-529	Sequence 529, App
24	33.2	6.7	2896	4 US-09-710-279-529	Sequence 529, App
25	33.2	6.7	4179	4 US-09-710-279-349	Sequence 349, Ap
26	33.2	6.7	81001	4 US-09-750-580-1	Sequence 1, Appl1
27	33	6.7	1410	4 US-09-861-451A-13	Sequence 13, Appl1

28	32.8	6.7	84495	3 US-09-797-906-3	Sequence 3, Appl1
29	32	6.5	202	4 US-09-513-999C-16073	Sequence 16073, A
30	32	6.5	372	3 US-08-991-789A-26	Sequence 26, Appl
31	32	6.5	372	3 US-09-062-451-26	Sequence 26, Appl
32	32	6.5	372	4 US-09-598-326-26	Sequence 26, Appl
33	32	6.5	372	4 US-09-598-326-26	Sequence 26, Appl
34	32	6.5	372	4 US-09-289-198-26	Sequence 26, Appl
35	32	6.5	2291	4 US-09-429-755-26	Sequence 26, Appl
36	32	6.5	161652	4 US-09-023-655-698	Sequence 698, Appl
37	31.8	6.5	7705	2 US-09-497-855A-40	Sequence 40, Appl
38	31.6	6.4	443	4 US-08-687-080-115	Sequence 115, App
39	31.6	6.4	443	4 US-09-702-705-1339	Sequence 1339, App
40	31.6	6.4	443	4 US-09-736-457-1339	Sequence 1339, App
41	31.6	6.4	443	4 US-09-614-124B-1339	Sequence 1339, App
42	31.6	6.4	443	4 US-09-671-325-1339	Sequence 1339, App
43	31.6	6.4	444	4 US-09-658-824-1339	Sequence 1339, App
44	31.6	6.4	444	4 US-09-702-705-1460	Sequence 1460, App
45	31.6	6.4	444	4 US-09-736-457-1460	Sequence 1460, App
				4 US-09-614-124B-1460	Sequence 1460, App

#### ALIGNMENTS

```
RESULT 1
US-09-744-128-3
; Sequence 3, Application US/09744128
; Patent No. 6677306
; GENERAL INFORMATION:
; APPLICANT: Vels et al.
; TITLE OF INVENTION: Chondrogenic and Osteogenic Inducing Molecule
; FILE REFERENCE: 27636/36983
; CURRENT APPLICATION NUMBER: US/09/744,128
; PRIOR APPLICATION NUMBER: PCT/US99/17342
; PRIOR FILING DATE: 1999-07-29
; PRIOR APPLICATION NUMBER: 60/094,489
; PRIOR FILING DATE: 1998-07-29
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Patentln 3.1
; SEQ ID NO 3
; LENGTH: 476
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: PCR Product
US-09-744-128-3
Query Match 10.4%; Score 51.2; DB 4; Length 476;
Best Local Similarity 94.6%; Pred. No. 1.1e-06;
Matches 53; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 ATGGGAGCCTGATTTGTTGCTGCTGCGAGACCTTGTATGCCGT 56
Db 48 ATGGGAGCCTGATCTGTTGCTGCTGCGAGACCTTGTATGCCGT 103
RESULT 2
US-09-018-595B-3
; Sequence 3, Application US/09018595B
; Patent No. 596223
; GENERAL INFORMATION:
; APPLICANT: Perkin-Elmer Corporation,
; APPLICANT: Applied Biosystems Division
; TITLE OF INVENTION: DETERMINATION OF GENOTYPE OF
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David J. Weitz,
; ADDRESSEE: Wilson Sonsini Goodrich & Rosati
; STREET: 650 Page Mill Road
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
```

```

ZIP: 34304-1050
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: Microsoft Windows 95
SOFTWARE: Wordperfect for windows 6.0,
SOFTWARE: ASCII (DOS) TEXT format
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/018,595B
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: David J. Weitz
REGISTRATION NUMBER: 38,362
REFERENCE/DOCKET NUMBER: PELM-744
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 493-9300
TELEFAX: (650) 493-6811
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 500 nucleotides
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
IS-09-018-595B-3

```

Query Match	9.8%	Score 48.2	DB 2	Length 500
Best Local Similarity	94.3%	Pred. NO. 1.1e-05		
Matches	50	Conservative	0	Mismatches 3, Indels 0, Gaps 0
QY	1	ATGGGAACTCGATGTTTGTTCCTGCTCCTCTGGAGAGACGCTTTGCTATGCC	53	
	69	ATGGGAACTCGATGTTTGTTCCTGCTCCTCTGGAGAGACGCTTTGCTATGCC	121	.

RESULT 3  
 US-09-324-709A-3  
 Sequence 3 Application US/09324709A  
 Patent No. 6154707  
 GENERAL INFORMATION:  
 APPLICANT: Perkin-Elmer Corporation,  
 APPLICANT: Applied Biosystems Division  
 TITLE OF INVENTION: FLUORESCENCE GENOTYPING AT MULTIPLE ALLELIC  
 TITLE OF INVENTION: SITES  
 NUMBER OF SEQUENCES: 11  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: David J. Weitz,  
 ADDRESSEE: Wilson Sonsini Goodrich & Rosati  
 STREET: 650 Page Mill Road  
 CITY: Palo Alto  
 STATE: California  
 COUNTRY: USA  
 ZIP: 94304-1050  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: 3.5 inch diskette  
 COMPUTER: IBM compatible  
 OPERATING SYSTEM: Microsoft Windows 95  
 SOFTWARE: Wordperfect for windows 6.0,  
 SOFTWARE: ASCII (DOS) TEXT format  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/324,709A  
 FILING DATE:  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER:  
 FILING DATE:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: David J. Weitz  
 REGISTRATION NUMBER: 38,362  
 REFERENCE/DOCKET NUMBER: 16442-758

```

: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (650) 493-9300
:
: TELEFAX: (650) 493-6811
:
: INFORMATION FOR SEQ ID NO: 3:
:
: SEQUENCE CHARACTERISTICS:
:
: LENGTH: 500 nucleotides
:
: TYPE: nucleic acid
:
: STRANDEDNESS: double
:
: TOPOLOGY: linear
:
: US-09-324-709A-3

```

Query Match 9.8%; Score 48.2; DB 3; Length 500;  
 Best Local Similarity 94.3%; Pred. No. 1.1e-05;  
 Matches 50; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

RESULT 4
US-09-018-595B-1
Sequence 1, Application US/09018595B
Patent No. 596223
GENERAL INFORMATION:
APPLICANT: Perkin-Elmer Corporation,
APPLICANT: Applied Biosystems Division
TITLE OF INVENTION: DETERMINATION OF GENOTYPE OF
TITLE OF INVENTION: AMPLIFICATION PRODUCTS AT MULTIPLE ALLELIC SITES
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: David J. Weltz,
ADDRESSEE: Wilson Sonsini Goodrich & Rosati
STREET: 650 Page Mill Road
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1050
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: Microsoft Windows 95
SOFTWARE: Wordperfect for windows 6.0,
SOFTWARE: ASCII (DOS) TEXT format
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/018,595B
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: David J. Weltz
REGISTRATION NUMBER: 38,362
REFERENCE/DOCKET NUMBER: PELM-744
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 493-9300
TELEFAX: (650) 493-6811
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 793 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-018-595B-1

Query Match          9.8%; Score 48.2; DB 2; Length 793;
Best Local Similarity 94.3%; Pred. No. 1.4e-05;
Matches   50; Conservative    0; Mismatches   3; Indels   0; Gaps   0;

CY      1 ATGGGAGCCGTGATTTTGTCTGCCTCGACAGACCCTTGCTATGCC   53
        |||||||
DB      69 ATGGGAGCCGTGATTTTATTTTCCTGCCTCTCGGAGCACACTTTTGCAATGCC   121

```

```

1  COUNTRY: USA
2  ZIP: 94304-1050
3  COMPUTER READABLE FORM:
4  MEDIUM TYPE: 3.5 inch diskette
5  COMPUTER: IBM compatible
6  OPERATING SYSTEM: Microsoft Windows 95
7  SOFTWARE: Wordperfect for windows 6.0,
8  SOFTWARE: ASCII (DOS) TEXT format
9  CURRENT APPLICATION DATA:
10 APPLICATION NUMBER: US/09/018,595B
11 FILING DATE:
12 CLASSIFICATION: 435
13 PRIOR APPLICATION DATA:
14 APPLICATION NUMBER:
15 ATTORNEY/AGENT INFORMATION:
16 NAME: David J. Weitz
17 REGISTRATION NUMBER: 38,362
18 REFERENCE/DOCKET NUMBER: PELM-744
19 TELECOMMUNICATION INFORMATION:
20 TELEPHONE: (650) 493-9300
21 TELEFAX: (650) 493-6811
22 INFORMATION FOR SEQ ID NO: 2:
23 SEQUENCE CHARACTERISTICS:
24 LENGTH: 802 nucleotides
25 TYPE: nucleic acid
26 STRANDEDNESS: single
27 TOPOLOGY: linear
28 US-09-018-595B-2
29
30 Query Match 9.5%; Score 46.6; DB 2; Length 802;
31 Best Local Similarly 92.5%; Pred. No. 4,56-05;
32 Matches 49; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
33
34 Oy 1 ATGGGAGACCTGGATTGTGTTGCTGCTCCTGGAGACGCTTGATAGCC 53
35 Db 69 ATGGGAGACCTGGATTGTGTTGCTGCTGCTTGGAGAGCGCTTTCGATGCC 121
36
37 RESULT 7
38 US-09-324-709A-2
39 Sequence 2, Application US/09324709A
40 Patent No. 6154707
41 GENERAL INFORMATION:
42 APPLICANT: Perkin-Elmer Corporation,
43 APPLICANT: Applied Biosystems Division
44 TITLE OF INVENTION: FLUORESCENCE GENOTYPING AT MULTIPLE ALLELIC
45 TITLE OF INVENTION: SITES
46 NUMBER OF SEQUENCES: 11
47 CORRESPONDENCE ADDRESS:
48 ADDRESSEE: David J. Weitz,
49 ADDRESSEE: Wilson Sonsini Goodrich & Rosati
50 STREET: 650 Page Mill Road
51 CITY: Palo Alto.
52 STATE: California
53 COUNTRY: USA
54 ZIP: 94304-1050
55 COMPUTER READABLE FORM:
56 MEDIUM TYPE: 3.5 inch diskette
57 COMPUTER: IBM compatible
58 OPERATING SYSTEM: Microsoft Windows 95
59 SOFTWARE: Wordperfect for windows 6.0,
60 SOFTWARE: ASCII (DOS) TEXT format
61 CURRENT APPLICATION DATA:
62 APPLICATION NUMBER: US/09/324,709A
63 FILING DATE:
64 CLASSIFICATION:
65 PRIOR APPLICATION DATA:
66 APPLICATION NUMBER:
67 FILING DATE:
68 ATTORNEY/AGENT INFORMATION:
69 NAME: David J. Weitz
70 REGISTRATION NUMBER: 38,362

```

REFERENCE/DOCKET NUMBER: 16842-758  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 493-9300  
TELEFAX: (650) 493-6811  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 802 nucleotides  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-324-709A-2

Query Match 9.5%; Score 46.6; DB 3; Length 802;  
Best Local Similarity 92.5%; Pred. No. 4.5e-05;  
Matches 49; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ATGGGAGCAGTGAATTTGTTGCTGCTGCGAGAGACCTTGTATGCC 53  
Db 69 ATGGGAGCAGTGAATTTGTTGCTGCTGCGAGAGACCTTGTATGCC 121

RESULT 8  
US-09-806-708B-22/c  
Sequence 22, Application US/09806708B  
Patent No. 6784342  
GENERAL INFORMATION:  
APPLICANT: The University of British Columbia  
TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants  
FILE REFERENCE: 4810-58741  
CURRENT APPLICATION NUMBER: US/09/806,708B  
PRIORITY FILING DATE: 2001-04-03  
PRIORITY APPLICATION NUMBER: US 60/147,133  
PRIORITY FILING DATE: 1999-08-04  
NUMBER OF SEQ ID NOS: 23  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 22  
LENGTH: 1141  
TYPE: DNA  
ORGANISM: Artificial sequence  
FEATURE:  
NAME/KEY: promoter  
LOCATION: (1)..(1141)  
OTHER INFORMATION: consensus sequence of A.c., L.a., and B.n. PAB1 promoters  
US-09-806-708B-22

Query Match 8.5%; Score 41.6; DB 4; Length 1141;  
Best Local Similarity 7.7%; Pred. No. 0.0023;  
Matches 32; Conservative 189; Mismatches 194; Indels 0; Gaps 0;

Qy 48 TATCCCGTGAAGTAAATACCCCGCATATATTCATTTTCAACAAGTGGAAATTAAG 107  
Db 541 TTTNNMMWSGBVRMAGTMMRRHNNNNNTDRTYYMMWRMBRTTYYDSCKNKSMMRG 482  
Qy 108 TCTGCCCCACAGTGTAACTTTAGGTTTAAAGACGATCAAGATGATGTCCTCAAA 167  
Db 481 NNWAMMMWMAANNADAGMDHWTMTGNTMMBRAMMMMAACRRAYCCNNNNNNACV 422  
Qy 168 TGTCTGTGTGTTTAAAGAAACCTTGGAAGCTTGTATATAAAAAATATATTTCCAGA 227  
Db 421 WHKHKMRMTWKYMMKACNNNNNBKAMRYVAMMYSRDITNTDMMWTSMBWHMYTVD 362  
Qy 228 TGCCCTCCACCAAGTATGATTCAGTAAGCAGAGTGGGGGAGTGGCCAGGACTCGCA 287  
Db 361 YTTMRAMNNNNNNNNRBCKTTSMWMMMDHNTHTCTYGNTWGSAYBMAASMMWAGASNB 302  
Qy 288 TTTTAAAGACCTCAGAGATTCGTGAGACAAATTAATCTTAAATATCATCGCCA 347  
Db 301 VTYWCMRMTYMGKTNN 242  
Qy 348 TCTCTAGATGAGAGAACTTTTAAAGGAGCCCTTGAAAGCCCTTCAGAGAAAGTCTCG 407  
Db 241 WYEMMYGKHMBWRRABHRSNNMMWVCNNKRYTVSWHYAHMYBKBAVAVGNNNNWKD 182

Qy 408 AACAGCTTAGGCAATACCAAAATGCCAATTTCTCTAAACCACCAATTTCTA 462  
Db 181 RMAHHHMCATNNNNMMWYATYHHHMKKGAATNNKTBABRDDBAHVATYMYR 127

RESULT 9  
US-08-956-171E-4303/c  
Sequence 4303, Application US/08956171E  
Patent No. 6593114  
GENERAL INFORMATION:  
APPLICANT: Charles Kunsch  
Gil H. Choi  
Patrick S. Dillon  
Craig A. Rosen  
Steven C. Barash  
Michael R. Fannon  
TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences  
NUMBER OF SEQUENCES: 5256  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/956,171E  
FILING DATE: 20-Oct-1997  
CLASSIFICATION: <Unknown>  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: 60/009,861  
FILING DATE: January 5, 1996  
APPLICATION NUMBER: 08/781,986  
FILING DATE: January 3, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Mark J. Hyman  
REGISTRATION NUMBER: 46,789  
REFERENCE/DOCKET NUMBER: PB248P1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (240) 314-1224  
TELEFAX: (301) 309-8439  
INFORMATION FOR SEQ ID NO: 4303:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 374 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 4303:  
US-08-956-171E-4303

Query Match 7.3%; Score 35.8; DB 4; Length 374;  
Best Local Similarity 51.6%; Pred. No. 0.094;  
Matches 82; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

Qy 288 TTTTAAACAGCAGCTCAGAGATTCGTGAGACAAATTAATCTTGAATATCATCGCCA 347  
Db 240 TTGTAAACAACTAATGATCAACAATGCGAAGAAATTTGTGAAGATTAATACCC 181  
Qy 348 TCTCTAGATGAGAGAACTTTTGAAGGAGCCCTTGAAGCCCTCCAGAGAAAGTCTCG 407  
Db 180 GATTAGTTAAAGATGCGGTGACAGCGCCGACGTTTGGCAGCGATGAGCGCGCTAA 121  
Qy 408 AACAGCTTAGGCAATACCAAAATGCCAATTTTCTC 446  
Db 120 ACTAGCTAGATAGCTGTAATCGTGTACCAATGTTATC 82

RESULT 10



US-08-781-986A-4303/c  
; Sequence 4303, Application US/08781986A  
; Patent No. 6737248  
; GENERAL INFORMATION:  
; APPLICANT: Charles Kunach  
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences  
; NUMBER OF SEQUENCES: 5255  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/781,986A  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Benson, Bob  
; REGISTRATION NUMBER: 30,446  
; REFERENCE/DOCKET NUMBER: PB248PP  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 309-8504  
; TELEFAX: (301) 309-8512  
; INFORMATION FOR SEQ ID NO: 4303:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 374 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; US-08-781-986A-4303  
  
Query Match 7.3%; Score 35.8; DB 4; Length 374;  
Best Local Similarity 51.6%; Pred. No. 0.094;  
Matches 82; Conservative 0; Mismatches 77; Indels 0; Gaps 0;  
  
QY 288 TTTTAAACAGACCTTCGAGATTCGTGAGACATTAATTGTAATATCATCGCCCA 347  
DB 240 TTGTAACTAACTAATCAATGATCACTGCGCAAGAAATGTGAAGATAATACCCC 181  
QY 348 TCCTGATGAGAGAACTTTGAGAGGACCTTGAAAGCCTCCAGAGAAAGTCTCG 407  
DB 180 GATTAAAGTTAAAGATGCGTGTACAGCGCGCAGCTTTTGACGCGATTGACCGGCTAA 121  
  
QY 408 AACAGCTTAGCAATATCTACAAATAATGCCAATTTTCTC 446  
DB 120 ACTAGCTAAGATAGCTGTATCGTGTACCAATGTTATC 82  
  
RESULT 11  
US-08-956-171E-3662  
; Sequence 3662, Application US/08956171E  
; Patent No. 6593114  
; GENERAL INFORMATION:  
; APPLICANT: Charles Kunach  
; GIL H. Choi  
; Patrick S. Dillon  
; Craig A. Rosen  
; Steven C. Barash  
; Michael R. Fannon  
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences  
; NUMBER OF SEQUENCES: 5256  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/956,171E  
; FILING DATE: 20-Oct-1997  
; CLASSIFICATION: <unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/009,861  
; FILING DATE: January 5, 1996  
; APPLICATION NUMBER: 08/781,986  
; FILING DATE: January 3, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mark U. Hyman  
; REGISTRATION NUMBER: 46,789  
; REFERENCE/DOCKET NUMBER: PB248P1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (240) 314-1224  
; TELEFAX: (301) 309-8439  
; INFORMATION FOR SEQ ID NO: 3662:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 742 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; US-08-956-171E-3662  
  
Query Match 7.3%; Score 35.8; DB 4; Length 742;  
Best Local Similarity 51.6%; Pred. No. 0.14;  
Matches 82; Conservative 0; Mismatches 77; Indels 0; Gaps 0;  
  
QY 288 TTTTAAACAGACCTTCGAGATTCGTGAGACATTAATTGTAATATCATCGCCCA 347  
DB 408 TTGTAACTAACTAATCAATGATCACTGCGCAAGAAATGTGAAGATAATACCCC 467  
QY 348 TCCTGATGAGAGAACTTTGAGAGGACCTTGAAAGCCTCCAGAGAAAGTCTCG 407  
DB 468 GATTAAAGTTAAAGATGCGTGTACAGCGCGCAGCTTTTGACGCGATTGACCGGCTAA 527  
DB 528 ACTAGCTAAGATAGCTGTATCGTGTACCAATGTTATC 566  
  
RESULT 12  
US-08-781-986A-3662  
; Sequence 3662, Application US/08781986A  
; Patent No. 6737248  
; GENERAL INFORMATION:  
; APPLICANT: Charles Kunach  
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences  
; NUMBER OF SEQUENCES: 5255  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/781,986A  
FILING DATE: 435  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Benson, Bob  
REGISTRATION NUMBER: 30,446  
REFERENCE/DOCKET NUMBER: PB248PP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 3662:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 742 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-781-986A-3662

Query Match  
Best Local Similarity 7.3%; Score 35.8; DB 4; Length 742;  
Pred. No. 0.14; Mismatches 0; Gaps 0;  
Matches 82; Conservative 0; Indels 77; Gaps 0;

QY 288 TTTTACAGCAGCAGGAGATTCTGTGAGACATTAATTGTAATATCATCGCCCA 347  
DB 408 TTGTACAACTAATCAATGATGATCACTGCGAGAGAAATTTGTAATATTC 467  
QY 348 TCTCTAGATGAGAGAACTTTTGAAGGAGACCTTGAAGGCTCCAGAGAAATGCTCG 407  
DB 468 GATTAAGTTAAAGATGACGTGTACAAAGCGCGACGTTTGCAGCGATTAGCGCGCTAA 527

QY 408 AACAGCTTAGGCAATATCTCAAAAATGCGCAATTTTCTC 446  
DB 528 ACTAGCTAAGATGCTGTAACTGTGTACCAATGTTATC 566

RESULT 13  
US-08-956-171E-221/C  
Sequence 221, Application US/08956171E  
Patent No. 6593114  
GENERAL INFORMATION:  
APPLICANT: Charles Kunesh  
Gil H. Choi  
Patrick S. Dillon  
Craig A. Rosen  
Steven C. Barash  
Michael R. Fannom  
TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences  
NUMBER OF SEQUENCES: 5256  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/956,171E  
FILING DATE: 20-Oct-1997  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/009,861  
FILING DATE: January 5, 1996  
APPLICATION NUMBER: 08/781,986  
FILING DATE: January 3, 1997  
ATTORNEY/AGENT INFORMATION:

NAME: Mark J. Hyman  
REGISTRATION NUMBER: 46,789  
REFERENCE/DOCKET NUMBER: PB248P1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (240) 314-1224  
TELEFAX: (301) 309-8439  
INFORMATION FOR SEQ ID NO: 221:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 10758 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-956-171E-221

Query Match  
Best Local Similarity 7.3%; Score 35.8; DB 4; Length 10758;  
Pred. No. 0.58; Mismatches 77; Indels 0; Gaps 0;  
Matches 82; Conservative 0; Indels 77; Gaps 0;

QY 288 TTTTACAGCAGCAGGAGATTCTGTGAGACATTAATTGTAATATCATCGCCCA 347  
DB 2860 TTGTACAACTAATCAATGATGATCACTGCGAGAGAAATTTGTAATATTC 2801  
QY 348 TCTCTAGATGAGAGAACTTTTGAAGGAGACCTTGAAGGCTCCAGAGAAATGCTCG 407  
DB 2800 GATTAAGTTAAAGATGACGTGTACAAAGCGCGACGTTTGCAGCGATTAGCGCGCTAA 2741

QY 408 AACAGCTTAGGCAATATCTCAAAAATGCGCAATTTTCTC 446  
DB 2740 ACTAGCTAAGATGCTGTAACTGTGTACCAATGTTATC 2702

RESULT 14  
US-08-781-986A-221/C  
Sequence 221, Application US/08781986A  
Patent No. 6737248  
GENERAL INFORMATION:  
APPLICANT: Charles Kunesh  
TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences  
NUMBER OF SEQUENCES: 5255  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/781,986A  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Benson, Bob  
REGISTRATION NUMBER: 30,446  
REFERENCE/DOCKET NUMBER: PB248PP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 221:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 10758 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-781-986A-221

Query Match 7.3%; Score 35.8; DB 4; Length 10758;  
Best Local Similarity 51.6%; Pred. No. 0.58; 77; Indels 0; Gaps 0;  
Matches 82; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 288 TTTTAAACAGCACCCTCAGAGATTCTGTGAGACATTAATTGTAATATCATCGCCA 347  
DB 2860 TTGTAACAACTAATCAATGAATCACACCTGGCAGAAATGTGAAGATTAATACCC 2801  
QY 348 TCTCTAGATGAGGAACTTTTGAAGGACCTTTGAAGGCTCCAGAGAAAGTCTCG 407  
DB 2800 GATTAACTTAAAGATGACGTGTACAGCGCCGACGTTTTCAGACGATGAGCGCGCTAA 2741  
QY 408 AACAGCTTAGGCAATCTACTCAAAAATGCCAATTTTCTC 446  
DB 2740 ACTAGCTAAGATAGCTGTAACTCGTGTACCAATGTATC 2702

## RESULT 15

US-09-621-976-399/C  
Sequence 399, Application US/09621976  
Patent No. 6639063  
GENERAL INFORMATION:  
APPLICANT: Dumas Milne Edwards, J. B.  
APPLICANT: Jobert, S.  
APPLICANT: Giordano, J. Y.  
TITLE OF INVENTION: ESTs and Encoded Human Proteins.  
FILE REFERENCE: GENSET.054PR2  
CURRENT APPLICATION NUMBER: US/09/621,976  
NUMBER OF SEQ ID NOS: 2000-07-21  
SOFTWARE: Patent.pm  
SEQ ID NO 399  
LENGTH: 370  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 139..369  
NAME/KEY: misc\_feature  
LOCATION: 128  
OTHER INFORMATION: n=a, g, c or t  
US-09-621-976-399

## Query Match

7.1%; Score 35; DB 4; Length 370;  
Best Local Similarity 63.1%; Pred. No. 0.17; 31; Indels 0; Gaps 0;  
Matches 53; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 231 CTCACCCCAAGACTGATTCAAGTAGAGAGAGTGGGGAGTGCCAGACTGTCATTT 290  
DB 135 CACCCCNAAAATGGCTGCCCGAATCAACAAATTTCTGATGGGCCAGCACTTGTGTTT 76  
QY 291 TAAACAAGCACCCTGAGAGATTCTG 314  
DB 75 TAAACAAGCTCTTCAGGGGATTCTG 52

Search completed: February 8, 2005, 22:26:27  
Job time : 134.991 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 8, 2005, 16:37:21 ; Search time 662.969 Seconds  
(without alignments)  
4264.119 Million cell updates/sec

Title: US-10-754-437-22  
Perfect score: 492  
Sequence: 1 atggggagctgagcttctgtc.....aactctctcgcgcctcca 492

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4300275 seqs, 2872944193 residues  
Total number of hits satisfying chosen parameters: 8600550

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Published Applications NA:  
1: /cgn2\_6/prodata/2/pubpna/us07\_PUBCOMB.seq.\*  
2: /cgn2\_6/prodata/2/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/prodata/2/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/prodata/2/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/prodata/2/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/prodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/prodata/2/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/prodata/2/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/prodata/2/pubpna/US09A\_PUBCOMB.seq.\*  
10: /cgn2\_6/prodata/2/pubpna/US09B\_PUBCOMB.seq.\*  
11: /cgn2\_6/prodata/2/pubpna/US09C\_PUBCOMB.seq.\*  
12: /cgn2\_6/prodata/2/pubpna/US09\_NEW\_PUB.seq.\*  
13: /cgn2\_6/prodata/2/pubpna/US10\_PUBCOMB.seq.\*  
14: /cgn2\_6/prodata/2/pubpna/US10\_PUBCOMB.seq.\*  
15: /cgn2\_6/prodata/2/pubpna/US10\_PUBCOMB.seq.\*  
16: /cgn2\_6/prodata/2/pubpna/US10\_PUBCOMB.seq.\*  
17: /cgn2\_6/prodata/2/pubpna/US10\_PUBCOMB.seq.\*  
18: /cgn2\_6/prodata/2/pubpna/US10\_NEW\_PUB.seq.\*  
19: /cgn2\_6/prodata/2/pubpna/US60\_NEW\_PUB.seq.\*  
20: /cgn2\_6/prodata/2/pubpna/US60\_PUBCOMB.seq.\*  
21: /cgn2\_6/prodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	274.6	55.8	259202	18	US-10-723-860-1311 Sequence 1311, Ap
2	124.2	25.2	600	13	US-10-027-632-287050 Sequence 287050,
3	124.2	25.2	600	15	US-10-027-632-287050 Sequence 287050,
4	123	25.0	600	13	US-10-027-632-287049 Sequence 287049,
5	123	25.0	600	15	US-10-027-632-287049 Sequence 287049,
6	48.2	9.8	500	13	US-10-104-774-3 Sequence 3, Appl1
7	48.2	9.8	500	16	US-10-455-150-3 Sequence 3, Appl1
8	48.2	9.8	793	13	US-10-104-774-1 Sequence 1, Appl1
9	48.2	9.8	793	16	US-10-455-150-1 Sequence 1, Appl1
10	48.2	9.8	793	17	US-10-755-889-605 Sequence 605, App
11	46.6	9.5	802	13	US-10-104-774-2 Sequence 2, Appl1
12	46.6	9.5	802	16	US-10-455-150-2 Sequence 2, Appl1

13	46.6	9.5	852	18	US-10-723-860-5797	Sequence 5797, Ap
14	45.2	9.2	665	10	US-09-908-975-1137	Sequence 1137, Ap
15	41.4	8.4	294	9	US-10-425-115-24717	Sequence 24717, A
16	39.2	8.0	397658	18	US-09-813-320-3	Sequence 3, Appl
17	38.4	7.8	773	14	US-10-198-846-3870	Sequence 3870, Ap
18	38.2	7.8	798	13	US-10-027-632-10730	Sequence 10730, A
19	38.2	7.8	798	13	US-10-027-632-10731	Sequence 10731, A
20	38.2	7.8	798	15	US-10-027-632-10730	Sequence 10730, A
21	38.2	7.8	798	15	US-10-027-632-10731	Sequence 10731, A
22	37.6	7.6	585	13	US-10-027-632-238025	Sequence 238025
23	37.6	7.6	585	15	US-10-027-632-238025	Sequence 238025,
24	36.6	7.4	3673778	15	US-10-312-884-1	Sequence 1, Appl
25	36.2	7.4	435	13	US-10-027-632-56046	Sequence 56046, A
26	36.2	7.4	435	13	US-10-027-632-57109	Sequence 57109, A
27	36.2	7.4	435	13	US-10-027-632-57925	Sequence 57925, A
28	36.2	7.4	435	15	US-10-027-632-56046	Sequence 56046, A
29	36.2	7.4	435	15	US-10-027-632-57109	Sequence 57109, A
30	36.2	7.4	435	15	US-10-027-632-57925	Sequence 57925, A
31	36.2	7.4	595	15	US-10-029-386-7878	Sequence 7878, Ap
32	36.2	7.4	911	13	US-10-027-632-104584	Sequence 104584,
33	36.2	7.4	911	15	US-10-027-632-104584	Sequence 104584,
34	36.2	7.4	1450	18	US-10-425-115-42306	Sequence 42306, A
35	36.2	7.4	166910	16	US-10-292-337-11	Sequence 11, Appl
36	36.2	7.4	176930	17	US-10-741-601-5679	Sequence 5679, Ap
37	36	7.3	425	13	US-10-027-632-182973	Sequence 182973,
38	36	7.3	425	15	US-10-027-632-182973	Sequence 182973,
39	35.8	7.3	374	8	US-08-781-986A-4303	Sequence 4303, Ap
40	35.8	7.3	374	16	US-10-329-62A-4303	Sequence 4303, Ap
41	35.8	7.3	742	8	US-08-781-986A-3662	Sequence 3662, Ap
42	35.8	7.3	742	16	US-10-329-62A-3662	Sequence 3662, Ap
43	35.8	7.3	10758	8	US-08-781-986A-221	Sequence 221, App
44	35.8	7.3	10758	16	US-10-329-62A-221	Sequence 221, App
45	35.6	7.2	7536	15	US-10-252-157-159	Sequence 159, App

#### ALIGNMENTS

RESULT 1  
US-10-723-860-1311  
; Sequence 1311, Application US/10723860  
; Publication No. US20040253606A1  
; GENERAL INFORMATION:  
; APPLICANT: Aziz, Natsasha  
; APPLICANT: Ginsburg, Wendy M.  
; APPLICANT: Zlotnik, Albert  
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &  
; TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators  
; FILE REFERENCE: 05882.0193.NPUS01  
; CURRENT APPLICATION NUMBER: US/10/723, 860  
; CURRENT FILING DATE: 2003-11-26  
; PRIOR APPLICATION NUMBER: 60/429, 739  
; PRIOR FILING DATE: 2002-11-26  
; NUMBER OF SEQ ID NOS: 8393  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 1311  
; LENGTH: 259202  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-723-860-1311

Query Match 55.8%; Score 274.6; DB 18; Length 259202;  
Best Local Similarity 79.0%; Pred. No. 2, 2e-72;  
Matches 394; Conservative 0; Mismatches 89; Indels 16; Gaps 5;  
QY 1 ATGGGAGCTGAGTTTGTTCCTGCTCTGAGAGAGCCCTTGCTATGCCGAGT 60  
DB 68873 ATGGGAGCTGAGTTTGTTCCTGCTCTGAGAGAGCCCTTGCTATGCCGAGT 69932  
QY 61 AAAATACCCCT-----GCATATATTCATTTTACAGAGCTTGAATTAAGCTGCC 114  
DB 69933 AAAAAGCCCTTCATATGACATGTCATTTTACAACTTGCATTAATTAATTCGCTC 69992

```
OY 115 CACAGTGTGTAACCTTTAGGGTTTAAAGACATGACATGATGCTGCAAAATGCTCT 174
DB 69993 ATAGTTGGTGAAA--TTAGGGTTTAAACAGTATGATGATGCTTCTCATATGCTCT 70050
OY 175 GTGTTTAAAGAAACACTTGGAGAGCTTGTATATAAAAAATATATATCCAGATGCTCC 234
DB 70051 GGGTTGAAGAAACACTT-CAGAGACTTGTATTAAGATATATCTCAATATGCGCTAC 70109
OY 235 ACCCAAGACTGATTCAGTAGAGCAGAGTGGGGGAGATGCCAGAGCTTGCATTTTAC 234
DB 70110 CAAAAATTCGATTTGATGACAGCTGGGGCGGG-----GCCAGAGCTTGCATTTTAT 70163
OY 295 AAGCAGCTCAGAGATTCGTGAGAGACATTTAATCTGTAATATCATGCGCAATCTAG 354
DB 70164 AAGCAGCTCAGAGATTCGTGAGAGCTTGTGAGACTTGTATATATCAACCACTCTTAG 70223
OY 355 ATGAGAGAACTTTTAAAGGAGACCTTGAAGGCTTCAGAGAAAGTGTGCAACAGCT 414
DB 70224 ATGAGAGAGCTTTTGAAGGAGACCTTGAAGGCTTCAGAGAAAGTGTGTAACAGCT 70283
OY 415 TAGG-CAAAATCTCAAAAATGCCAATTTCTCTAAACCCTTTTAAAGAGTGTCCA 473
DB 70284 TTGGACAAATATTAACAAGATGCCAGTTTGTCTAAACCCTTTCTCAAGATTCCA 70343
OY 474 ACTCTCTTCTGCGCCCTCCA 492
DB 70344 AATCTCTTCTGCGCCCTCCA 70362
```

```
RESULT 2
US-10-027-632-287050/c
; Sequence 287050, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 287050
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-287050
```

```
Query Match 25.2%; Score 124.2; DB 13; Length 600;
Best Local Similarity 75.3%; Pred. No. 2e-27;
Matches 195; Conservative 1; Mismatches 54; Indels 9; Gaps 3;
OY 1 ATGGGAGCTGATTTTGTTCCTGCTCTGAGAGACCTTTGCTATGCCGTAGT 60
DB 256 ATGGGAGCTGATTTTATTTGCTGCTCTGAGAGACCTTTTGCATGCTGAGT 197
OY 61 AAAATACCTT-----GCATATATTCATTTTCAAGCTTGGAAATTAAGTGTGCC 114
DB 196 AAAACACCTTGCATTAAGTCAAGTGTCCAAATTTTCAAACTTGAGCATTAATAATCTGCYC 137
```

```
OY 115 CACAGTGTGTAACCTTTAGGGTTTAAAGACATGACATGATGCTGCAAAATGCTCT 174
DB 136 ATAGTTGGTGAAA--TTAGGGTTTAAACAGTATGATGATGCTTCTCATATGCTCT 79
OY 175 GTGTTTAAAGAAACACTTGGAGAGCTTGTATATAAAAAATATATATCCAGATGCTCC 234
DB 78 GGGTTGAAGAAACACTT-CAGAGACTTGTATTAAGATATATCTCAATATGCGCTAC 20
OY 235 ACCCAAGACTGATTCAGTA 253
DB 19 CAAAAATTCGATTTGTGA 1
```

```
RESULT 3
US-10-027-632-287050/c
; Sequence 287050, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 287050
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-287050
```

```
Query Match 25.2%; Score 124.2; DB 15; Length 600;
Best Local Similarity 75.3%; Pred. No. 2e-27;
Matches 195; Conservative 1; Mismatches 54; Indels 9; Gaps 3;
OY 1 ATGGGAGCTGATTTTGTTCCTGCTCTGAGAGACCTTTGCTATGCCGTAGT 60
DB 256 ATGGGAGCTGATTTTATTTGCTGCTCTGAGAGACCTTTTGCATGCTGAGT 197
OY 61 AAAATACCTT-----GCATATATTCATTTTCAAGCTTGGAAATTAAGTGTGCC 114
DB 196 AAAACACCTTGCATTAAGTCAAGTGTCCAAATTTTCAAACTTGAGCATTAATAATCTGCYC 137
OY 115 CACAGTGTGTAACCTTTAGGGTTTAAAGACATGACATGATGCTGCAAAATGCTCT 174
DB 136 ATAGTTGGTGAAA--TTAGGGTTTAAACAGTATGATGATGCTTCTCATATGCTCT 79
OY 175 GTGTTTAAAGAAACACTTGGAGAGCTTGTATATAAAAAATATATATCCAGATGCTCC 234
DB 78 GGGTTGAAGAAACACTT-CAGAGACTTGTATTAAGATATATCTCAATATGCGCTAC 20
OY 235 ACCCAAGACTGATTCAGTA 253
DB 19 CAAAAATTCGATTTGTGA 1
```

RESULT 4

```
US-10-027-632-287049/c
; Sequence 287049, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 287049
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-287049
```

```
Query Match      25.0%; Score 123; DB 13; Length 600;
Best Local Similarity 75.3%; Pred. No. 4,7e-27;
Matches 195; Conservative 0; Mismatches 55; Indels 9; Gaps 3;
```

```
QY 1 ATGGGACCTGATTTTGTGCTGCTGCTGAGACAGCTTTGCTATGCCGTAGT 60
    |||||
DB 256 ATGGGACCTGATTTTATTTGCTGCTGCTGAGACAGCTTTGCTATGCCGTAGT 197
    |||||
QY 61 AAATAACCCCT-----GCATATATTCATTTTCAAGCTTGGAAATAAAGCTGCC 114
    |||||
DB 196 AAAACACCCCTTGATAGTCAAGTCCAAATTCACAACTTGACATTAATAATCTGCTC 137
    |||||
QY 115 CACAGTTGTTAACTTTAGAGTTTAAAGACATCAAGATGATGCTCAATGCTCT 174
    |||||
DB 136 ATAGTTGGTAAA--TTAGGTTTAAACAGTATGATGATGATGCTCTCATATGCTCT 79
    |||||
QY 175 GTGTTTAAAGAACCTTGGAAGAGCTGTTATATAAAAAATATATTCACAGATGCTCC 234
    |||||
DB 78 GGGTTGAAGAACCTT-CAGGAGCTGTTTAAAAAGTATATTCACAAATGCGCTAC 20
    |||||
QY 235 ACCCAAGACTGATTCAGTA 253
    |||||
DB 19 CAAAAATTCGATTGGTA 1
    |||||
```

```
RESULT 5
US-10-027-632-287049/c
; Sequence 287049, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
```

```
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 287049
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-287049
```

```
Query Match      25.0%; Score 123; DB 15; Length 600;
Best Local Similarity 75.3%; Pred. No. 4,7e-27;
Matches 195; Conservative 0; Mismatches 55; Indels 9; Gaps 3;
```

```
QY 1 ATGGGACCTGATTTTGTGCTGCTGCTGAGACAGCTTTGCTATGCCGTAGT 60
    |||||
DB 256 ATGGGACCTGATTTTATTTGCTGCTGCTGAGACAGCTTTGCTATGCCGTAGT 197
    |||||
QY 61 AAATAACCCCT-----GCATATATTCATTTTCAAGCTTGGAAATAAAGCTGCC 114
    |||||
DB 196 AAAACACCCCTTGATAGTCAAGTCCAAATTCACAACTTGACATTAATAATCTGCTC 137
    |||||
QY 115 CACAGTTGTTAACTTTAGAGTTTAAAGACATCAAGATGATGCTCAATGCTCT 174
    |||||
DB 136 ATAGTTGGTAAA--TTAGGTTTAAACAGTATGATGATGATGCTCTCATATGCTCT 79
    |||||
QY 175 GTGTTTAAAGAACCTTGGAAGAGCTGTTATATAAAAAATATATTCACAGATGCTCC 234
    |||||
DB 78 GGGTTGAAGAACCTT-CAGGAGCTGTTTAAAAAGTATATTCACAAATGCGCTAC 20
    |||||
QY 235 ACCCAAGACTGATTCAGTA 253
    |||||
DB 19 CAAAAATTCGATTGGTA 1
    |||||
```

```
RESULT 6
US-10-104-774-3
; Sequence 3, Application US/10104774
; Publication No. US20020164630A1
; GENERAL INFORMATION:
; APPLICANT: Perkin-Elmer Corporation, Applied Biosystems Division
; TITLE OF INVENTION: DETERMINATION OF GENOTYPE OF AMPLIFICATION PRODUCTS AT MULTIPLE
; TITLE OF INVENTION: SITES
; FILE REFERENCE: 16842-782
; CURRENT APPLICATION NUMBER: US/10/104,774
; CURRENT FILING DATE: 2002-03-21
; PRIOR APPLICATION NUMBER: US 09/018,595
; PRIOR FILING DATE: 1998-02-04
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 3
; LENGTH: 500
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-104-774-3
```

```
Query Match      9.8%; Score 48.2; DB 13; Length 500;
Best Local Similarity 94.3%; Pred. No. 0.00045;
Matches 50; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 1 ATGGGACCTGATTTTGTGCTGCTGCTGAGACAGCTTTGCTATGCC 53
    |||||
DB 69 ATGGGACCTGATTTTATTTGCTGCTGCTGAGACAGCTTTGCTATGCC 121
    |||||
```

RESULT 7

US-10-455-150-3  
; Sequence 3, Application US/10455150  
; Publication No. US20040053302A1  
; GENERAL INFORMATION:  
; APPLICANT: Perkin-Elmer Corporation,  
; Applied Biosystems Division  
; TITLE OF INVENTION: FLUORESCENCE GENOTYPING AT MULTIPLE ALLELIC  
; SITES  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: David J. Weitz,  
; Wilson Sonsini Goodrich & Rosati  
; STREET: 650 Page Mill Road  
; CITY: Palo Alto  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94304-1050  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 inch diskette  
; COMPUTER: IBM compatible  
; OPERATING SYSTEM: Microsoft Windows 95  
; SOFTWARE: Wordperfect for windows 6.0,  
; ASCII (DOS) TEXT format  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/455,150  
; FILING DATE: 04-Jun-2003  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/09/324,709A  
; FILING DATE: 03-June-1999  
; ATTORNEY/AGENT INFORMATION:  
; NAME: David J. Weitz  
; REGISTRATION NUMBER: 38,362  
; REFERENCE/DOCKET NUMBER: 16842-758  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (650) 493-9300  
; TELEFAX: (650) 493-6811  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 500 nucleotides  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; SEQUENCE DESCRIPTION: SEQ ID NO: 3:  
US-10-455-150-3  
Query Match 9.8%; Score 48.2; DB 16; Length 500;  
Best Local Similarity 94.3%; Pred. No. 0.00045;  
Matches 50; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 ATGGGACCTGGATTGTTGCTGCTGCTGCGAGCAGCTTTGCTATGCC 53  
Db 69 ATGGGACCTGGATTGTTGCTGCTGCTGCGAGCAGCTTTGCTATGCC 121  
RESULT 8  
US-10-104-774-1  
; Sequence 1, Application US/10104774  
; Publication No. US20020164630A1  
; GENERAL INFORMATION:  
; APPLICANT: Perkin-Elmer Corporation, Applied Biosystems Division  
; TITLE OF INVENTION: DETERMINATION OF GENOTYPE OF AMPLIFICATION PRODUCTS AT MULTIPLE A  
; SITES  
; FILE REFERENCE: 16842-782  
; CURRENT APPLICATION NUMBER: US/10/104,774  
; CURRENT FILING DATE: 2002-03-21  
; PRIOR APPLICATION NUMBER: US 09/018,595  
; PRIOR FILING DATE: 1998-02-04  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 1  
; LENGTH: 793  
; TYPE: DNA

ORGANISM: Homo sapiens  
US-10-104-774-1  
Query Match 9.8%; Score 48.2; DB 13; Length 793;  
Best Local Similarity 94.3%; Pred. No. 0.00056;  
Matches 50; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 ATGGGACCTGGATTGTTGCTGCTGCTGCGAGCAGCTTTGCTATGCC 53  
Db 69 ATGGGACCTGGATTGTTGCTGCTGCTGCGAGCAGCTTTGCTATGCC 121  
RESULT 9  
US-10-455-150-1  
; Sequence 1, Application US/10455150  
; Publication No. US20040053302A1  
; GENERAL INFORMATION:  
; APPLICANT: Perkin-Elmer Corporation,  
; Applied Biosystems Division  
; TITLE OF INVENTION: FLUORESCENCE GENOTYPING AT MULTIPLE ALLELIC  
; SITES  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: David J. Weitz,  
; Wilson Sonsini Goodrich & Rosati  
; STREET: 650 Page Mill Road  
; CITY: Palo Alto  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94304-1050  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 inch diskette  
; COMPUTER: IBM compatible  
; OPERATING SYSTEM: Microsoft Windows 95  
; SOFTWARE: Wordperfect for windows 6.0,  
; ASCII (DOS) TEXT format  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/455,150  
; FILING DATE: 04-Jun-2003  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/09/324,709A  
; FILING DATE: 03-June-1999  
; ATTORNEY/AGENT INFORMATION:  
; NAME: David J. Weitz  
; REGISTRATION NUMBER: 38,362  
; REFERENCE/DOCKET NUMBER: 16842-758  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (650) 493-9300  
; TELEFAX: (650) 493-6811  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 793 nucleotides  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-10-455-150-1  
Query Match 9.8%; Score 48.2; DB 16; Length 793;  
Best Local Similarity 94.3%; Pred. No. 0.00056;  
Matches 50; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 ATGGGACCTGGATTGTTGCTGCTGCTGCGAGCAGCTTTGCTATGCC 53  
Db 69 ATGGGACCTGGATTGTTGCTGCTGCTGCGAGCAGCTTTGCTATGCC 121  
RESULT 10  
US-10-755-889-605  
; Sequence 605, Application US/10755889  
; Publication No. US20040171823A1  
; GENERAL INFORMATION:



```
APPLICANT: Bristol-Myers Squibb Company
TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ASSOCIATED WITH THE NF-KB
FILE REFERENCE: D0284 NP
CURRENT APPLICATION NUMBER: US/10/755,889
CURRENT FILING DATE: 2004-01-13
PRIOR APPLICATION NUMBER: U.S. 60/440,068
PRIOR FILING DATE: 2003-01-14
PRIOR APPLICATION NUMBER: U.S. 60/469,757
PRIOR FILING DATE: 2003-05-12
NUMBER OF SEQ ID NOS: 823
SOFTWARE: PatentIn version 3.2
SEQ ID NO 605
LENGTH: 793
TYPE: DNA
ORGANISM: Homo sapiens
US-10-755-889-605

Query Match
Best Local Similarity 9.8%; Score 48.2; DB 17; Length 793;
Matches 50; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATGGGACCTGATTTGTTGCTGCTGCTGCGAGACAGCTTGTGCTATGCC 53
DB 69 ATGGGACCTGATTTGTTGCTGCTGCTGCGAGACAGCTTGTGCTATGCC 121

RESULT 11
US-10-104-774-2
Sequence 2, Application US/10104774
Publication No. US20020164630A1
GENERAL INFORMATION:
APPLICANT: Perkin-Elmer Corporation, Applied Biosystems Division
TITLE OF INVENTION: DETERMINATION OF GENOTYPE OF AMPLIFICATION PRODUCTS AT MULTIPLE A
FILE REFERENCE: 16842-782
CURRENT APPLICATION NUMBER: US/10/104,774
CURRENT FILING DATE: 2002-03-21
PRIOR APPLICATION NUMBER: US 09/018,595
PRIOR FILING DATE: 1998-02-04
NUMBER OF SEQ ID NOS: 11
SOFTWARE: PatentIn version 3.1
SEQ ID NO 2
LENGTH: 802
TYPE: DNA
ORGANISM: Homo sapiens
US-10-104-774-2

Query Match
Best Local Similarity 9.5%; Score 46.6; DB 13; Length 802;
Matches 49; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ATGGGACCTGATTTGTTGCTGCTGCTGCGAGACAGCTTGTGCTATGCC 53
DB 69 ATGGGACCTGATTTGTTGCTGCTGCTGCGAGACAGCTTGTGCTATGCC 121

RESULT 12
US-10-455-150-2
Sequence 2, Application US/10455150
Publication No. US20040053302A1
GENERAL INFORMATION:
APPLICANT: Perkin-Elmer Corporation,
TITLE OF INVENTION: FLUORESCENCE GENOTYPING AT MULTIPLE ALLELIC
SITES
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: David J. Weitz,
Walson Sonsini Goodrich & Rosati
STREET: 650 Page Mill Road
CITY: Palo Alto
STATE: California
```

```
COUNTRY: USA
ZIP: 94304-1050
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: Microsoft Windows 95
SOFTWARE: Wordperfect for windows 6.0,
ASCII (DOS) TEXT format

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/455,150
FILING DATE: 04-Jun-2003
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/324,709A
FILING DATE: 03-June-1999
ATTORNEY/AGENT INFORMATION:
NAME: David J. Weitz
REGISTRATION NUMBER: 38,362
REFERENCE/DOCKET NUMBER: 16842-758
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 493-9300
TELEFAX: (650) 493-6811
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 802 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-10-455-150-2

Query Match
Best Local Similarity 9.5%; Score 46.6; DB 16; Length 802;
Matches 49; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ATGGGACCTGATTTGTTGCTGCTGCTGCGAGACAGCTTGTGCTATGCC 53
DB 69 ATGGGACCTGATTTGTTGCTGCTGCTGCGAGACAGCTTGTGCTATGCC 121

RESULT 13
US-10-723-860-5797
Sequence 5797, Application US/10723860
Publication No. US20040253606A1
GENERAL INFORMATION:
APPLICANT: Aziz, Natasha
APPLICANT: Ginsburg, Wendy M.
TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
FILE REFERENCE: 05882.0193.NPUS01
CURRENT APPLICATION NUMBER: US/10/723,860
CURRENT FILING DATE: 2003-11-26
PRIOR APPLICATION NUMBER: 60/429,739
PRIOR FILING DATE: 2002-11-26
NUMBER OF SEQ ID NOS: 8393
SOFTWARE: PatentIn version 3.2
SEQ ID NO 5797
LENGTH: 852
TYPE: DNA
ORGANISM: Homo sapiens
US-10-723-860-5797

Query Match
Best Local Similarity 9.5%; Score 46.6; DB 18; Length 852;
Matches 49; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ATGGGACCTGATTTGTTGCTGCTGCTGCGAGACAGCTTGTGCTATGCC 53
DB 69 ATGGGACCTGATTTGTTGCTGCTGCTGCGAGACAGCTTGTGCTATGCC 121

RESULT 14
```

US-09-908-975-1137  
; Sequence 1137, Application US/09908975  
; Publication No. US20030165843A1  
; GENERAL INFORMATION:  
; APPLICANT: SHOSHAN, Avi  
; APPLICANT: WASSERMAN, Alon  
; APPLICANT: MINTZ, Eli  
; APPLICANT: MINTZ, Ilat  
; APPLICANT: FAIGLER, Simchon  
; TITLE OF INVENTION: OLIGONUCLEOTIDE LIBRARY FOR DETECTING RNA TRANSCRIPTS AND SPLICING  
; FILE REFERENCE: 36688-0005  
; CURRENT APPLICATION NUMBER: US/09/908,975  
; PRIOR FILING DATE: 2001-07-20  
; PRIOR APPLICATION NUMBER: US 60/267,724  
; PRIOR FILING DATE: 2001-05-02  
; PRIOR APPLICATION NUMBER: US 60/221,607  
; PRIOR FILING DATE: 2000-07-28  
; NUMBER OF SEQ ID NOS: 32337  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 1137  
; LENGTH: 65  
; TYPE: DNA  
; ORGANISM: Rattus norvegicus  
US-09-908-975-1137

Query Match 9.2%; Score 45.2; DB 10; Length 65;  
Best Local Similarity 94.0%; Pred. No. 0.0014;  
Matches 47; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 7 ACCTGATTTTGTGTTGCTGCTGCTGAGACACCTTTGCTATGCCCGT 56  
DB 1 ACCTGATCTTGTGTTGCTGCTGCTGAGACACCTTTGCTATGCCCGT 50

RESULT 15  
US-10-425-115-24717  
; Sequence 24717, Application US/10425115  
; Publication No. US20040214272A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yinhua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants  
; FILE REFERENCE: 38-21 (53222)B  
; CURRENT APPLICATION NUMBER: US/10/425,115  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 369326  
; SEQ ID NO 24717  
; LENGTH: 294  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURES:  
; OTHER INFORMATION: Clone ID: MRT4577\_122549C.1  
US-10-425-115-24717

Query Match 8.4%; Score 41.4; DB 18; Length 294;  
Best Local Similarity 54.2%; Pred. No. 0.042;  
Matches 84; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 58 AGTAAATACCTGCAATATATTCATTTACAGCTTGGAATATAAGTCTGCCCCAC 117  
DB 112 AGTCACTAGTCTGTTACATTTGATTCACCTGCTCAAGCAAAATGGATCCAA 171  
QY 118 AGTTGTAACCTTTAGGCTTTAAGACAGTCAAGATCAAGTCTCAATGCTCTGTG 177  
DB 172 TGTCGCGCATTCGTGTAAACGTATTAATCTTTAACTGCTCAACTATCAATTTT 231  
QY 178 TTTAAGAACACTTGAAGAGCTTGTATTAATAA 212  
DB 232 TTTATATATAATGCTTGTGTTAAATAAATAA 266

Search completed: February 9, 2005, 06:51:20  
Job time : 667.969 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using bw model

Run on: February 8, 2005, 12:58:25 ; Search time 5660.18 Seconds  
(without alignments)  
3059.349 Million cell updates/sec

Title: US-10-754-437-22

Perfect score: 492  
Sequence: 1 atggggagactgagcttctgtt.....aactctctctgcctcca 492

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
EST:  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_hnc:\*  
4: gb\_est3:\*  
5: gb\_est4:\*  
6: gb\_est5:\*  
7: gb\_est6:\*  
8: gb\_g881:\*  
9: gb\_g882:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysts of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	130.2	26.5	516	8	AQ003910	CIT-HSP-2
C 2	107.6	21.9	544	8	BH118642	RPCI-24-3
C 3	98.6	20.0	513	8	A2725508	RPCI-24-1
4	81.6	16.6	790	8	BH080441	RPCI-24-3
5	52.8	10.7	303	7	W33932	mb54d02.r1
6	52.8	10.7	430	7	W40649	mc43h12.r1
7	52.8	10.7	536	7	W36345	mb72c12.r1
8	52.8	10.7	585	9	AY119441	mus muscu
9	52.8	10.7	645	6	CD773419	AGENCOURT
10	52.8	10.7	707	6	CB056709	NISC_j119
11	52.8	10.7	812	3	AK029358	mus muscu
12	52.8	10.7	843	6	CB588212	AGENCOURT
13	52.8	10.7	859	6	CB588525	AGENCOURT
14	52.8	10.7	865	6	CB589177	AGENCOURT
15	52.8	10.7	886	6	CB587051	AGENCOURT
16	52.8	10.7	889	6	CB589251	AGENCOURT
17	52.8	10.7	891	6	CB574837	AGENCOURT
18	52.8	10.7	913	6	CB590451	AGENCOURT
19	52.8	10.7	919	6	CB587332	AGENCOURT
20	52.8	10.7	928	6	CB590111	AGENCOURT
21	52.8	10.7	945	2	BB614068	BB614068
22	52.8	10.7	990	7	W12906	ma89p03.r1
23	52.8	10.7	1006	7	W29475	mb99f11.r1
24	52.8	10.7	1020	7	W08102	mb40g08.r1

25	51.2	10.4	217	7	R46913	R46913 Y140 Rat in
26	51.2	10.4	223	2	BB571643	BB571643 BB571643
27	51.2	10.4	449	7	R47024	R47024 Y350 Rat in
28	51.2	10.4	1031	6	CB575508	CB575508 AGENCOURT
29	50.2	10.2	373	7	R47135	R47135 Y722 Rat in
30	50.2	10.2	395	7	R47030	R47030 Y359 Rat in
31	50.2	10.2	556	7	R46934	R46934 Y169 Rat in
32	49.6	10.1	257	7	R47100	R47100 Y601 Rat in
33	49.2	10.0	303	7	R46903	R46903 Y124 Rat in
34	49.2	10.0	318	7	R47078	R47078 Y534 Rat in
35	49.2	10.0	467	7	R46947	R46947 Y186 Rat in
36	48.6	9.9	293	7	R47143	R47143 Y79 Rat inc
37	48.2	9.8	429	6	CB473334	CB473334 sm68 B03
38	48.2	9.8	570	9	AY119439	AY119439 Homo_sapi
39	48.2	9.8	570	9	AY119440	AY119440 Pan trogl
40	48.2	9.8	395	7	R46955	R46955 Y195 Rat in
41	47.6	9.7	478	7	R46933	R46933 Y166 Rat in
42	47.2	9.6	273	7	R47002	R47002 Y3 Rat incl
43	46.6	9.5	286	7	R46887	R46887 Y106 Rat in
44	46.6	9.5	783	4	BG198114	BG198114 RST17499
45	46.6	9.3	355	7	R46940	R46940 Y174 Rat in

## ALIGNMENTS

RESULT 1  
AQ003910/c 516 bp DNA linear GSS 26-JUN-1998  
CIT-HSP-2289L4.TF CIT-HSP Homo sapiens genomic clone 2289L4,  
DEFINITION  
genomic survey sequence.  
ACCESSION  
AQ003910 GI:3081561  
VERSION  
AQ003910.1 GI:3081561  
KEYWORDS  
GSS.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
1 (bases 1 to 516)  
Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Baas,S., Linher,K.,  
Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,  
Simon,M. and Venter,J.C.  
Use of a random BAC End Sequence Database for Sequence-Ready Map  
Building (1998)  
JOURNAL  
Unpublished (1998)  
COMMENT  
Other GSSs: CIT-HSP-2289L4.TF  
Contact: Mark Adams  
Department of Bukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: mdadams@igf.org  
Clones are available from Research Genetics (info@resgen.com). BAC  
end search page:  
http://www.igf.org/cdb/human/bac\_end\_search/bac\_end\_search.html  
Seq primer: M13-21;  
Class: BAC ends.

## FEATURES

### source

1. 516  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="GDB:7150375"  
/db\_xref="taxon:9606"  
/clone="2289L4"  
/sex="Male"  
/cell\_type="Sperm"  
/clone\_1lb="CIT-HSP"  
/note="Vector: pBelobAC11, Site\_1: HindIII, Site\_2:  
HindIII"

## ORIGIN

Query Match 26.5%; Score 130.2; DB 8; Length 516;

Best Local Similarity 82.6%; Pred. No. 8e-27; Matches 161; Conservative 0; Mismatches 33; Indels 1; Gaps 1;

QY 299 ACCCTGAGAGATTCGTGGAGACATTAATTGTAATATCATGCCCATCTAGATGG 358  
 Db 516 ACCCTGAGAGATTCGTGGAGACATTAATTGTAATATCATGCCCATCTAGATGG 457  
 QY 359 AGGAACTTTGAGAGAGACCTTGAAGGCTCCAGAGAAAGTCTGAAGAGCTTAGG 418  
 Db 456 AGAAGCTTTGAGAGAGACCTTGAAGGCTCCAGAGAAAGTCTGAAGAGCTTAGG 397  
 QY 419 -CAATATCTAGAAAATATCCCAATTTCTCTAAAACCAATTTCTAGAGTGTCAATC 477  
 Db 396 ACAATATCTAGAGATGCCAGTTTGTCTAAAACCAATTTCTCTAGATTTCAATC 337  
 QY 478 TCTTCTGCTCTCCA 492  
 Db 336 TCTTCTGCTCTCCA 322

RESULT 2  
 BHL18642/c 544 bp DNA linear GSS 19-JUN-2001  
 LOCUS RPCI-24-359A2.TV RPCI-24 Mus musculus genomic clone RPCI-24-359A2,  
 DEFINITION genomic survey sequence.

ACCESSION BHL18642  
 VERSION BHL18642.1 GI:14961134  
 KEYWORDS GSS.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 544)  
 AUTHORS Zhao S., Nierman, W., Malek, J., Shatsman, S., Akintet, B., Levins, M., Teegate, G., Geer, K., Krol, M., Shvartsbeyn, A., Gebregorgis, E., Russell, D., de Jong, P. and Fraser, C.M.  
 JOURNAL Mouse BAC End Sequences from Library RPCI-24  
 COMMENT Unpublished (1999)  
 Other GSSes: RPCI-24-359A2.TV  
 Contact: Shaying Zhao  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-24. For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end page: http://www.tigr.org/tdb/bac\_ends/mouse/bac\_end\_intro.html  
 Plate: 359 row: A column: 2  
 Seq primer: SP6  
 Class: BAC ends

# FEATURES

Source location/Qualifiers  
 1. 344  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="RPCI-24-359A2"  
 /sex="Male"  
 /cell\_type="Spleen/Brain"  
 /clone\_lib="RPCI-24"  
 /notes="Vector: PTARBAC1; Site 1: BamHI; Site 2: BamHI; RPCI-24 Mouse BAC library produced by Pieter de Jong. The library was cloned in the pTARBAC1 cloning vector at the BamHI sites using MboI partially digested male C57BL/6J DNA."

ORIGIN  
 Query Match 21.9%; Score 107.6; DB 8; Length 544;  
 Best Local Similarity 61.6%; Pred. No. 3.3e-20;

Matches 261; Conservative 0; Mismatches 149; Indels 14; Gaps 5;

QY 77 ATATTCATTTTCAAGCTTGAATAATGAAGTGG-CCACAGTTGGTAACTTAAAGG 135  
 Db 496 ATCCCATTTGAATTAATCTTGAACCTGAATATGCTCCCATAGTTGTAATTTCACTG 437  
 QY 136 TTTAAGACGTACAGATCATGATGCTCCAAATGCTGTGTTTAAAGAAACATTTGAA 195  
 Db 436 TGTAAACAGTGCAGATCTTAATATCCCAAGATTAATAGGTGTTGAAGAAATCATTT 377  
 QY 196 GAGCTGTATATAAAAAAATATATCCAGATGCTCCACCAAC-TCATTCAGTA 253  
 Db 376 TTAGTTTGTATTAAGAA-ATACAGAAAGTCTCACTTACAGCTGTGATTAAGTT 321  
 QY 254 GAGCAGAGTGGGGAGAGGCCAGAGACTGTGATTTTAAACAAGCACTCAGAGATTTCT 313  
 Db 320 CAGCTGAAGTGGGACCAAGACATCAAGATTTTAAACAAGATTTGACCTCAGAGATCTT 261  
 QY 314 GTGAGACATTAATCTTGAATATATATGATCCCATCTCTAGATGAGAACTTTAGAA 373  
 Db 260 CTTGACAGATTAAGCTTGTAAACACACAGCTCCCATCTTACAAAGTATATTTGAA 201  
 QY 374 GGGACCTTGAAGAGGCTCCAGAGAAAGTCTGAACAGCTTAGGCAATATACAAATA 433  
 Db 200 GGGACCCAGAAAGAGCTTCCAGAG-AGTGTACTTGTGTTTAAAGCAATATTAACAGCCC 143  
 QY 434 TGCC---AATTTCTTAAACCATTTTGAAGAGTGTCAACTCTTCTGCCCC 488  
 Db 142 CCCCCCAAAAAAATCTTAAGCTTAATTAATTAATTAATTAATTAATTTCTTGTGCC 83  
 QY 489 TCCA 492  
 Db 82 TCCA 79

RESULT 3  
 AZ725508/c 513 bp DNA linear GSS 24-JAN-2001  
 LOCUS RPCI-24-10901.TV RPCI-24 Mus musculus genomic clone RPCI-24-10901,  
 DEFINITION genomic survey sequence.

ACCESSION AZ725508  
 VERSION AZ725508.1 GI:12472231  
 KEYWORDS GSS.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 513)  
 AUTHORS Zhao S., Nierman, W., Malek, J., Shatsman, S., Akintet, B., Levins, M., Teegate, G., Geer, K., Krol, M., Shvartsbeyn, A., Gebregorgis, E., Russell, D., de Jong, P. and Fraser, C.M.  
 JOURNAL Mouse BAC End Sequences from Library RPCI-24  
 COMMENT Unpublished (1999)  
 Contact: Shaying Zhao  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-24. For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end page: http://www.tigr.org/tdb/bac\_ends/mouse/bac\_end\_intro.html  
 Plate: 109 row: J column: 1  
 Seq primer: T7  
 Class: BAC ends

# FEATURES

Source location/Qualifiers  
 1. 513  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /strain="C57BL/6J"

/db xref="taxon:10090"  
/clone="RPCI-24-10931"  
/sex="Male"  
/cell\_type="Spleen/Brain"  
/clone\_lib="RPCI-24"  
/notes="Vector: pTARBAcl; Site 1: BamHI; Site 2: BamHI;  
RPCI-24 Mouse BAC library produced by Pieter de Jong. The  
library was cloned in the pTARBAcl cloning vector at the  
BamHI sites using MboI partially digested male C57BL/6J  
DNA."

## ORIGIN

Query Match 20.4%; Score 98.6; DB 8; Length 513;  
Best Local Similarity 60.1%; Pred. No. 1.4e-17;  
Matches 255; Conservative 0; Mismatches 156; Indels 13; Gaps 5;

77 ATATTCATTTCACAGCTTGAATAATTAAGTCTGC-CCCAAGTTGGTAACTTTGAGG 135  
475 ATCCCAATTGAATATCTTGAACCTGAATATCTGCCATGTTGTAATTTTCACTG 416  
136 TTTTAAGACAGTACAGATCAGATGCTCTCAATGCTCTGTTTAAAGAACTTGGAA 195  
415 TGTAAACAGCTGCAAGATCTTAATATCCCAAGTATCTAATAGGTGTAAGAAATCATTT 356  
196 GAGCTGTATTAATAAAAAATATTTCCAGATGCTCCCAAGAC--TGAATTCAGTA 253  
355 TTGAGTTTGTAAAGAA-ATATCAGAAAGTCTCCAGCTTGAAGCTGATTCAGTT 300  
254 GAGCAGAGTGGGGAGTCCCAAGACCTGCTGCTTTTAAAGACCTCAGAGATCTT 313  
299 CAGCTGAAGTGGGCAAGACCTCAAGATTTTAAAGACCTTGAAGCTTGAAGATCTT 240  
314 GTGAGACATTAATCTTGAATATTCATCGCCCTCTCTGATGAGAGAACTTTAGAA 373  
239 CTGTAAAGTATGCTTGTAAACACCTGCCATCTCTCAAGTATTTATTTTGA 180  
374 GGGACCTTGAAGGCTCCAGAGAAAGTGTGAACAGCTTGAAGCAATATCAAAAA 433  
179 GAGAACTTGAAGGCTTTC-CAAGAGTGTATCTTGTATGAGCAATATCAAGCCC 121  
434 TGCC-----AATTTCTTAAACCAATTTCTAAGAGTGTCAACTCTCTTCTGCC 488  
120 CCCCCCAAAAAATCTTAAGCTTAATTAAGTGTCAAGATTTCCAAATTTCTTCTGCC 61  
489 TCCTCA 492  
60 TCCTCA 57

RESULT 4  
LOCUS BH080441 790 bp DNA linear GSS 18-JUL-2001  
DEFINITION RPCI-24-321F13, TJ RPCI-24 Mus musculus genomic clone  
ACCESSION BH080441  
VERSION BH080441.1 GI:14900038  
KEYWORDS GSS.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 790)  
Zhaio,S., Nierman,W., Malek,J., Shatsman,S., Akincir,B., Levins,M.,  
Teegye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregorgis,E.,  
Ruesell,D., de Jong,P. and Fraser,C.M.  
Mouse BAC End Sequences from Library RPCI-24  
Unpublished (1999)  
Contact: Shaying Zhao  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208

TITLE JOURNAL  
COMMENT

Email: szhao@tigr.org  
Clones are derived from the mouse BAC library RPCI-24. For BAC  
library availability, please contact Pieter de Jong  
(pdejong@mail.cho.org). Clones may be purchased from BACPAC  
Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end  
page: [http://www.tigr.org/tdb/bac\\_ends/mouse/bac\\_end\\_intro.html](http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html)  
Plate: 321 row: F column: 13  
Seq primer: SP6  
Class: BAC ends.

## FEATURES

## Source

Location/Qualifiers  
1..790

/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone\_lib="RPCI-24-321F13"  
/sex="Male"  
/cell\_type="Spleen/Brain"  
/clone\_lib="RPCI-24"  
/notes="Vector: pTARBAcl; Site 1: BamHI; Site 2: BamHI;  
RPCI-24 Mouse BAC library produced by Pieter de Jong. The  
library was cloned in the pTARBAcl cloning vector at the  
BamHI sites using MboI partially digested male C57BL/6J  
DNA."

## ORIGIN

Query Match 16.4%; Score 81.6; DB 8; Length 790;  
Best Local Similarity 62.4%; Pred. No. 1.5e-12;  
Matches 181; Conservative 0; Mismatches 99; Indels 10; Gaps 3;

211 AAAATATATNTCCAGATGCTCCACCAAGAC--TGATTCATAGCAGAGTGGGG 268  
29 AAAAGAAATCAGAAAGCTCCACCTTGAAGTATGATTCATGAGTGGAG 88  
269 GAGTGCCAGAGCTGCAATTTAAGACAGCTCAGAGATTTCTGAGACAAATTAAC 328  
89 CAGACCTCAAGTTTAAAGACATTTGACCCCTGAGATCTTCTGTAAGATTAGC 148  
329 TTGTAATATCATGCCCATCTCTAAGTGAAGAACTTTAAGAGGACCTTGAAGG 388  
149 TTGTAACAACACTGCCCATCTCTAAGATGTTATTTGAAGGACCTTGAAGG 208  
389 CTTCCAGAAAGTGTCCAGACAGCTTGAAGCAATCTAAGAAATGCCC-----AATTT 442  
209 CTTCCAGAG--AGTGTATCTTGTATGAGCAATATTAACAGCCCCCCCCCAAAAA 266  
443 TCTTAAACCAATTTCTAAGAGTGTCCAACTCTTCTGCTCCCTCCA 492  
267 TCTTAAAGCTTAATTAAGTGTCCAAATTTCTTCTGCTCCCA 316

RESULT 5  
LOCUS W33932 303 bp mRNA linear EST 11-SEP-1996  
DEFINITION mb54d02.r1 Soares mouse p3NM19.5 Mus musculus cDNA clone  
IMAGE:333219 5' similar to gb:U10095 Mouse amelogenin (MOUSE);  
IMAGE:333219 5' similar to gb:U10095 Mouse amelogenin (MOUSE);  
W33932  
W33932.1 GI:115837  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 303)  
Maria,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
Schellendberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
Waterston,R.  
The WashU-HMI Mouse EST Project  
Unpublished (1996)  
Contact: Maria W/Mouse EST Project

TITLE JOURNAL  
COMMENT

WashU-HMI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@watsn.wustl.edu  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI:214619  
 Seq primer: mob.BEGA+ET  
 High quality sequence stop: 185.  
 Location/Qualifiers  
 1..303  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:333219"  
 /dev\_stage="19.5 dpc total fetus"  
 /lab\_host="DH10B (ampicillin resistant)"  
 /clone\_lib="Soares mouse p3NM19.5"  
 /note="Vector: pT73D (Pharmacia) with a modified  
 polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
 was primed with a Not I - oligo(dT) primer [5',  
 TGTACCAATCTGAAGTGGAGCGCGCATTTTCTTTTCTTTT 3']  
 double-stranded cDNA was size selected, ligated to Eco RI  
 adapters (Pharmacia), digested with Not I and cloned into  
 the Not I and Eco RI sites of a modified pT73 vector  
 (Pharmacia). Library went through one round of  
 normalization to a Cot = 5. Library constructed by Bento  
 Soares and M. Fatima Bonaldo. RNA was kindly provided by  
 Dr. Minoru Ko (Wayne State University)."

## ORIGIN

Query Match 10.7%; Score 52.8; DB 7; Length 303;  
 Best Local Similarity 96.4%; Pred. No. 0.00032;  
 Matches 54; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGGGACCTGATTTGTTGCTGCTCTCGGAGACGCTTGTATGCCCGT 56  
 |||||  
 Db 70 ATGGGACCTGATTTGTTGCTGCTCTCGGAGACGCTTGTATGCCCGT 125

RESULT 6  
 W40649  
 LOCUS  
 DEFINITION  
 430 bp mRNA linear EST 11-SEP-1996  
 m443b12.r1 Soares mouse p3NM19.5 Mus musculus cDNA clone  
 IMAGE:351335 5' similar to gb:M10095 Mouse amelogenin (MOUSE);,  
 mRNA sequence.  
 W40649  
 VERSION  
 W40649.1 GI:1324982  
 EST.  
 SOURCE  
 Mus musculus (house mouse)  
 ORGANISM  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 430)  
 REFERENCE  
 1 (bases 1 to 430)  
 Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
 Geisels,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
 Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
 Theisinger,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
 Waterston,R.  
 TITLE  
 The WashU-HMI Mouse EST Project  
 JOURNAL  
 Unpublished (1996)  
 COMMENT  
 Contact: Maria M/Mouse EST Project  
 WashU-HMI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@watsn.wustl.edu  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI:223135

Seq primer: ETPprimer  
 High quality sequence stop: 323.  
 Location/Qualifiers  
 1..430  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:351335"  
 /dev\_stage="19.5 dpc total fetus"  
 /lab\_host="DH10B (ampicillin resistant)"  
 /clone\_lib="Soares mouse p3NM19.5"  
 /note="Vector: pT73D (Pharmacia) with a modified  
 polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
 was primed with a Not I - oligo(dT) primer [5',  
 TGTACCAATCTGAAGTGGAGCGCGCATTTTCTTTTCTTTT 3']  
 double-stranded cDNA was size selected, ligated to Eco RI  
 adapters (Pharmacia), digested with Not I and cloned into  
 the Not I and Eco RI sites of a modified pT73 vector  
 (Pharmacia). Library went through one round of  
 normalization to a Cot = 5. Library constructed by Bento  
 Soares and M. Fatima Bonaldo. RNA was kindly provided by  
 Dr. Minoru Ko (Wayne State University)."

## ORIGIN

Query Match 10.7%; Score 52.8; DB 7; Length 430;  
 Best Local Similarity 96.4%; Pred. No. 0.00035;  
 Matches 54; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGGGACCTGATTTGTTGCTGCTCTCGGAGACGCTTGTATGCCCGT 56  
 |||||  
 Db 58 ATGGGACCTGATTTGTTGCTGCTCTCGGAGACGCTTGTATGCCCGT 113

RESULT 7  
 W36345  
 LOCUS  
 DEFINITION  
 536 bp mRNA linear EST 11-SEP-1996  
 mb72c12.r1 Soares mouse p3NM19.5 Mus musculus cDNA clone  
 IMAGE:334966 5' similar to gb:M10095 Mouse amelogenin (MOUSE);,  
 mRNA sequence.  
 W36345  
 VERSION  
 W36345.1 GI:1318120  
 EST.  
 SOURCE  
 Mus musculus (house mouse)  
 ORGANISM  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 536)  
 REFERENCE  
 1 (bases 1 to 536)  
 Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
 Geisels,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
 Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
 Theisinger,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
 Waterston,R.  
 TITLE  
 The WashU-HMI Mouse EST Project  
 JOURNAL  
 Unpublished (1996)  
 COMMENT  
 Contact: Maria M/Mouse EST Project  
 WashU-HMI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@watsn.wustl.edu  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI:216366  
 Seq primer: ETPprimer  
 High quality sequence stop: 359.  
 Location/Qualifiers  
 1..536  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:334966"  
 /dev\_stage="19.5 dpc total fetus"





REFERENCE 1 (bases 1 to 707)  
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 JOURNAL Tumor Gene Index  
 COMMENT Unpublished (1997)  
 Contact: Robert Strausberg, Ph.D.  
 Email: [cgapdb-remail.nih.gov](mailto:cgapdb-remail.nih.gov)  
 CDNA Library Preparation:  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium/LNL  
 DNA Sequencing by: National Institutes of Health Intramural  
 Sequencing Center (NISC)  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.lnl.gov>  
 Plate: LLM10779 row: N column: 9  
 Seq primer: T7 primer.  
 Location/Qualifiers  
 1..707  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:4848584"  
 /cdate="1997-01-15"  
 /dev\_stage="juvenile, 13-15 days"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: brain; Vector: pTZ19-3; Site 1: NotI;  
 Site 2: EcoRI; 1st strand cDNA was primed with a NotI-  
 oligo(dT) primer  
 5'-AATCGGACCTGATTTGTTGCTGCTGCTGAGACAGCTTTGCTATGCCCGT  
 double-stranded cDNA was ligated to EcoRI adaptors  
 5'-AATCGGACGAGG-3' and 5'-CTCGGCGG-3' (Pharmacia),  
 digested with NotI and cloned into the NotI and EcoRI  
 sites of the pTZ19-3 vector. Library went through one  
 round of normalization, and was constructed in the  
 laboratory of M. Bento Soares (University of Iowa)."

ORIGIN  
 Query Match 10.7%; Score 52.8; DB 6; Length 707;  
 Best Local Similarity 96.4%; Pred. No. 0.00039;  
 Matches 54; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 47 ATGGGACCTGATTTGTTGCTGCTGCTGAGACAGCTTTGCTATGCCCGT 102

QY  
 1 ATGGGACCTGATTTGTTGCTGCTGCTGAGACAGCTTTGCTATGCCCGT 56  
 47 ATGGGACCTGATTTGTTGCTGCTGCTGAGACAGCTTTGCTATGCCCGT 102

Db  
 47 ATGGGACCTGATTTGTTGCTGCTGCTGAGACAGCTTTGCTATGCCCGT 102

RESULT 11  
 AK029358 812 bp mRNA linear HTC 03-APR-2004  
 LOCUS Mus musculus 0 day neonate head cDNA, RIKEN full-length enriched  
 library, clone:4833404E21 product:amelogenin, full insert sequence.  
 ACCESSION AK029358  
 VERSION AK029358.1 GI:26325321  
 KEYWORDS HTC; CAP trapper.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryotes; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
 AUTHORS Carninci, P. and Hayashizaki, Y.  
 TITLE High-efficiency full-length cDNA cloning  
 JOURNAL Meth. Enzymol. 303, 19-44 (1999)  
 MEDLINE 98279253  
 PUBMED 10349636

REFERENCE  
 AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
 Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.  
 TITLE Normalization and subtraction of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new genes  
 JOURNAL Genome Res. 10 (10), 1617-1630 (2000)  
 MEDLINE 20493374  
 PUBMED 11042159

REFERENCE 3  
 AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagao, S., Sasaki, N., Carninci, P.,  
 Kono, H., Akiyama, J., Nishi, K., Katsunari, T., Hashino, H., Itoh, M.,  
 Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Hazada, A.,  
 Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,  
 Fujisawa, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M.,  
 Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,  
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.  
 RIKEN integrated sequence analysis (RISA) system-384-format  
 sequencing pipeline with 384 multicapillary sequencer  
 Genome Res. 10 (11), 1757-1771 (2000)  
 MEDLINE 20530913  
 PUBMED 11076861

REFERENCE 4  
 AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the  
 PANTOM Consortium.  
 TITLE Functional annotation of a full-length mouse cDNA collection  
 Nature 409, 685-690 (2001)  
 MEDLINE 11976861

REFERENCE 5  
 AUTHORS The PANTOM Consortium and the RIKEN Genome Exploration Research  
 Group Phase I & II Team.  
 TITLE Analysis of the mouse transcriptome based on functional annotation  
 of 60,770 full-length cDNAs  
 Nature 420, 563-573 (2002)  
 MEDLINE 11976861

REFERENCE 6  
 AUTHORS (Bases 1 to 812)  
 FADACH, J., Aizawa, K., Akiyama, T., Arakawa, T., Bono, H., Carninci, P.,  
 Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,  
 Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T.,  
 Horii, P., Imoto, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,  
 Kato, H., Kawai, J., Kojima, Y., Kondo, S., Kono, H., Kouda, M.,  
 Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,  
 Nakamura, M., Nishi, K., Nomura, K., Nunataki, R., Ohno, M., Ohsato, N.,  
 Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,  
 Sano, H., Sasaki, D., Shibata, K., Shingawa, A., Shiraki, T.,  
 Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akanishi, S.,  
 Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,  
 Muramatsu, M., and Hayashizaki, Y.

COMMENT  
 Direct Submission  
 Submitted (16-JUN-2001) Yoshihide Hayashizaki, The Institute of  
 Physical and Chemical Research (RIKEN), Laboratory for Genome  
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
 RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,  
 Kanagawa 230-0045, Japan (E-mail: genome-res@gscc.riken.jp,  
 URL: <http://genome.gsc.riken.jp/>, Tel: 81-45-503-9222,  
 Fax: 81-45-503-9216)  
 CDNA library was prepared and sequenced in Mouse Genome  
 Encyclopedia Project of Genome Exploration Research Group in Riken  
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
 Division of Experimental Animal Research in Riken contributed to  
 prepare mouse tissues.  
 Please visit our web site for further details.  
 URL: <http://genome.gsc.riken.jp/>  
 URL: <http://fantom.gsc.riken.jp/>  
 Location/Qualifiers  
 1..812  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="FANTOM\_DB:4833404E21"  
 /db\_xref="taxon:10090"  
 /clone="4833404E21"  
 /cdate="1997-01-15"  
 /dev\_stage="0 day neonate"  
 /dev\_stage="0 day neonate"  
 /dev\_stage="0 day neonate"  
 /note="unnamed protein product; amelogenin (MGD) [GI:88005,  
 GB|D31768, evidence: BLASTN, 99%, match=799]  
 putative"  
 /protein\_id="BAC26415.1"  
 /db\_xref="GI:26325322"  
 /translation="MGTWLFACLLGAAPFAPRLPAPGPGYINLSYEVLTPLKMYOS  
 MRPQPSGYEPMGWLHNIQIIVLSQOHPPSHYLPQHNLIPVPAQOPVAPQOPPM



PVGGHSMPTQHHOPINIPSAOQFOPOAIPPOSHQPMQPOSPLHMQPLAVQ  
PRLPLFSMQPLSPILPELPLKAMPATDKKREBVD"

polya\_signal  
793..798  
/note="putative"  
polya\_site  
812  
/note="putative"

## ORIGIN

Query Match 10.7%; Score 52.8; DB 3; Length 812;  
Best Local Similarity 96.4%; Pred. No. 0.00041;  
Matches 54; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGGGAGCTGATTTGTTGCTGCTGCTGAGAGACCTTGTATGCCCGT 56  
|||||  
73 ATGGGAGCTGATTTGTTGCTGCTGCTGAGAGACCTTGTATGCCCGT 128

RESULT 12  
CB588212 843 bp mRNA linear EST 03-APR-2003  
LOCUS  
DEFINITION  
AGENCOURT\_12771359 NIH\_MGC\_136 Mus musculus cDNA clone  
IMAGE:30293239 5', mRNA sequence.  
CB588212  
CB588212.1 GI:29506068  
EST.  
KEYWORDS  
SOURCE  
ORGANISM  
Mus musculus (house mouse)

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 843)  
NIH-MGC http://mgi.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: Dr. David Rowe  
CDNA Library Preparation: Invitrogen Corp  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
http://image.llnl.gov  
Plate: NDAM338 row: 9 column: 08  
High quality sequence stop: 636.  
Location/Qualifiers

## FEATURES

source

1..843  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:30293239"  
/issue\_type="embryonic limb, maxilla and mandible"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_136"  
/note="Vector: pCMV-SPORT6.1; Site 1: EcoRV; Site 2: NotI;  
Normalized, full-length enriched library from pool of  
mouse embryonic limb, maxilla and mandible, embryonic day  
17.5, 18.5 and newborn (mandible (5, 4 and 1 limb and jaw  
equivalents from respective days). Cloned directionally,  
oligo-dt primed (5'-GACTAGTTTAAATGCGAGCGCCGCC(T)15-3'.  
Size selected for the >1kb fragments, average insert size  
1.2 kb. Normalization to Cot 7.5. Tissue contributed by  
David Rowe; library constructed by Resgen, Invitrogen  
Corp. Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 10.7%; Score 52.8; DB 6; Length 843;  
Best Local Similarity 96.4%; Pred. No. 0.00041;  
Matches 54; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGGGAGCTGATTTGTTGCTGCTGCTGAGAGACCTTGTATGCCCGT 56  
|||||  
40 ATGGGAGCTGATTTGTTGCTGCTGCTGAGAGACCTTGTATGCCCGT 95

## RESULT 13

CB588525 859 bp mRNA linear EST 03-APR-2003  
LOCUS

DEFINITION  
AGENCOURT\_12567944 NIH\_MGC\_136 Mus musculus cDNA clone  
IMAGE:30289751 5', mRNA sequence.  
CB588525

ACCESSION  
CB588525.1 GI:29506381  
EST.  
KEYWORDS  
SOURCE  
ORGANISM  
Mus musculus (house mouse)

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 859)  
NIH-MGC http://mgi.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: Dr. David Rowe  
CDNA Library Preparation: Invitrogen Corp  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
http://image.llnl.gov  
Plate: NDAM329 row: e column: 24  
High quality sequence stop: 763.  
Location/Qualifiers

## FEATURES

source

1..859  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:30289751"  
/issue\_type="embryonic limb, maxilla and mandible"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_136"  
/note="Vector: pCMV-SPORT6.1; Site 1: EcoRV; Site 2: NotI;  
Normalized, full-length enriched library from pool of  
mouse embryonic limb, maxilla and mandible, embryonic day  
17.5, 18.5 and newborn (mandible (5, 4 and 1 limb and jaw  
equivalents from respective days). Cloned directionally,  
oligo-dt primed (5'-GACTAGTTTAAATGCGAGCGCCGCC(T)15-3'.  
Size selected for the >1kb fragments, average insert size  
1.2 kb. Normalization to Cot 7.5. Tissue contributed by  
David Rowe; library constructed by Resgen, Invitrogen  
Corp. Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 10.7%; Score 52.8; DB 6; Length 859;  
Best Local Similarity 96.4%; Pred. No. 0.00041;  
Matches 54; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGGGAGCTGATTTGTTGCTGCTGCTGAGAGACCTTGTATGCCCGT 56  
|||||  
59 ATGGGAGCTGATTTGTTGCTGCTGCTGAGAGACCTTGTATGCCCGT 114

RESULT 14  
CB589177 865 bp mRNA linear EST 03-APR-2003  
LOCUS  
DEFINITION  
AGENCOURT\_12770446 NIH\_MGC\_136 Mus musculus cDNA clone  
IMAGE:30290534 5', mRNA sequence.  
CB589177

ACCESSION  
CB589177.1 GI:29507033  
EST.  
KEYWORDS  
SOURCE  
ORGANISM  
Mus musculus (house mouse)

REFERENCE  
AUTHORS  
TITLE  
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 865)  
NIH-MGC http://mgi.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)

## JOURNAL COMMENT

Unpublished (1999)  
 Contact: Robert Straubeberg, Ph.D.  
 Email: cgabs-remail.nih.gov  
 Tissue Procurement: Dr. David Rowe  
 CDNA Library Preparation: Invitrogen Corp  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 CDNA distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.llnl.gov>  
 Plate: NDAM31 row: f column: 15  
 High quality sequence stop: 630.

## FEATURES

## Source

Location/Qualifiers  
 1..865  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:30290534"  
 /tissue\_type="embryonic limb, maxilla and mandible"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_1b="NIH\_MGC\_136"  
 /note="Vector: PCMV-SPORT6.1; Site 1: EcoRV; Site 2: NotI; Normalized, full-length enriched library from pool of mouse embryonic limb, maxilla and mandible, embryonic day 17.5, 18.5 and newborn (mandible (5, 4 and 1 limb and jaw equivalents from respective days). Cloned directionally, oligo-dt primed (5'-GACTGATCTAGATCGGAGCGCGCC(T)15-3'. Size selected for the >1kb fragments, average insert size 1.2 kb. Normalization to Cot 7.5. Tissue contributed by David Rowe; library constructed by Resgen, Invitrogen Corp. Note: this is a NIH\_MGC library."

## ORIGIN

Query Match 10.7%; Score 52.8; DB 6; Length 865;  
 Best Local Similarity 96.4%; Pred. No. 0.00041;  
 Matches 54; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGGGACCTGGATTGTTGCTGCTGCTGCGAGCAGCTTGTCTATGCCCGT 56  
 |||||||  
 DB 58 ATGGGACCTGGATTGTTGCTGCTGCTGCGAGCAGCTTGTCTATGCCCGT 113

## RESULT 15

## CB587051

## LOCUS

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

CB587051 886 bp mRNA linear EST 03-APR-2003  
 AGENCOURT\_12807074 NIH\_MGC\_136 Mus musculus CDNA clone  
 IMAGE:30295123 5', mRNA sequence.  
 CB587051  
 CB587051.1 GI:29504907  
 EST.  
 Mus musculus (house mouse)  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 886)  
 NIH-MGC <http://mgs.nci.nih.gov/>.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Straubeberg, Ph.D.  
 Email: cgabs-remail.nih.gov  
 Tissue Procurement: Dr. David Rowe  
 CDNA Library Preparation: Invitrogen Corp  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 CDNA distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.llnl.gov>  
 Plate: NDAM343 row: e column: 20  
 High quality sequence stop: 619.  
 Location/Qualifiers  
 1..886  
 /organism="Mus musculus"  
 /mol\_type="mRNA"

## FEATURES

## Source

Location/Qualifiers  
 1..886  
 /organism="Mus musculus"  
 /mol\_type="mRNA"

## ORIGIN

Query Match 10.7%; Score 52.8; DB 6; Length 886;  
 Best Local Similarity 96.4%; Pred. No. 0.00042;  
 Matches 54; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGGGACCTGGATTGTTGCTGCTGCTGCGAGCAGCTTGTCTATGCCCGT 56  
 |||||||  
 DB 84 ATGGGACCTGGATTGTTGCTGCTGCTGCGAGCAGCTTGTCTATGCCCGT 139

Search completed: February 8, 2005, 22:18:21  
 Job time : 5861.18 secs

/db\_xref="taxon:10090"  
 /clone="IMAGE:30295123"  
 /tissue\_type="embryonic limb, maxilla and mandible"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_1b="NIH\_MGC\_136"  
 /note="Vector: PCMV-SPORT6.1; Site 1: EcoRV; Site 2: NotI; Normalized, full-length enriched library from pool of mouse embryonic limb, maxilla and mandible, embryonic day 17.5, 18.5 and newborn (mandible (5, 4 and 1 limb and jaw equivalents from respective days). Cloned directionally, oligo-dt primed (5'-GACTGATCTAGATCGGAGCGCGCC(T)15-3'. Size selected for the >1kb fragments, average insert size 1.2 kb. Normalization to Cot 7.5. Tissue contributed by David Rowe; library constructed by Resgen, Invitrogen Corp. Note: this is a NIH\_MGC library."

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 8, 2005, 11:28:00 ; Search time 3791.46 Seconds  
(without alignments)  
8481.433 Million cell updates/sec

Title: US-10-754-437-23

Perfect score: 680  
1 atcgggagactgagcttctgtc.....ggtactgagagacagacac 680

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

1: gb\_ba:\*  
2: gb\_ptg:\*  
3: gb\_in:\*  
4: gb\_ov:\*  
5: gb\_ov:\*  
6: gb\_pac:\*  
7: gb\_ph:\*  
8: gb\_dl:\*  
9: gb\_pt:\*  
10: gb\_ro:\*  
11: gb\_gte:\*  
12: gb\_gy:\*  
13: gb\_un:\*  
14: gb\_vl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	220	32.4	5712	4	AB091793 Equus cab
2	184.2	27.1	6451	4	AB091789 Bos tauru
3	181.2	26.6	7425	4	AB091791 Sus scrofa
4	162	23.8	6465	9	AB091781 Pan trogl
5	160.4	23.6	8810	9	AY040206 Homo sapi
6	160.4	23.6	158142	2	AL357130 Homo sapi
7	160.4	23.6	259202	2	AC002366 Human Xp2
8	145.2	21.4	5684	9	AB091785 Lemur cat
9	139.6	20.5	6442	9	AB091783 Saimiri s
10	135.4	19.9	5591	4	AB091794 Equus cab
11	129.8	19.1	6264	4	AB091790 Bos tauru
12	123.8	18.2	5562	9	AB091787 Odocoileu
13	118.4	17.4	1935	9	D83730 Homo sapien
14	95.2	14.0	7454	9	AB091784 Saimiri s
15	84	12.4	8004	9	AB091786 Lemur cat
16	81.4	12.0	2501	9	D83729 Homo sapien
17	81.4	12.0	7163	9	AB091782 Pan trogl
18	81.4	12.0	38765	9	BS000568 Pan trogl
19	81.4	12.0	17726	9	AC145770 Pan trogl

C	20	81.4	12.0	190089	9	BS000576 Pan trogl
C	21	81.4	12.0	200214	9	AC013412 Homo sapi
C	22	78.2	11.5	9384	10	AF294357 Mus muscu
C	23	78.2	11.5	95826	10	AL805974 Mouse DNA
C	24	73.4	10.8	212886	2	AC093946 Rattus no
C	25	73.4	10.8	234471	2	AC121424 Rattus no
C	26	72.8	10.7	695	11	BV089295 RPAMSE00
C	27	72.8	10.7	695	11	BV097603 RPAMSE00
C	28	69.2	10.2	6931	4	AB091792 Sus scrofa
C	29	67.2	9.9	463	10	D83064 Mus muscu
C	30	60.2	8.9	363	11	BV089294 RPAMSE00
C	31	60.2	8.9	363	11	BV097602 RPAMSE00
C	32	54.8	8.1	177654	2	AP000918 Homo sapi
C	33	51.6	7.6	142000	9	AP003110 Homo sapi
C	34	51.2	7.5	752	4	AF215890 Capra hie
C	35	51.2	7.5	752	4	AF215889 Capra hie
C	36	51.2	7.5	757	4	BOVAMLGNI1
C	37	51.2	7.5	820	4	BOVAMLGNI1
C	38	50.8	7.5	72832	8	AC074228 Bovine (cla
C	39	49.6	7.3	778	4	SSU43405 Arabidops
C	40	47.2	6.9	175630	2	AC010925 Sus scrofa
C	41	46.6	6.9	296	10	AC010925 Sus scrofa
C	42	46.6	6.9	727	10	MUSMAMB
C	43	46.6	6.9	765	4	AB032194 Equus cab
C	44	46.6	6.9	789	4	AB032193 Equus cab
C	45	46.6	6.9	798	10	BC059090 Mus muscu

## ALIGNMENTS

RESULT 1  
AB091793 5712 bp DNA linear MAM 02-MAY-2003  
LOCUS Equus caballus AMELX gene for amelogenin, partial cds.  
DEFINITION AB091793  
ACCESSION AB091793  
VERSION AB091793.1 GI:29126038

KEYWORDS  
SOURCE  
ORGANISM  
EQUUS caballus (horse)

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
12672962  
2 (bases 1 to 5712)  
Iwase,M., Saita,Y., Hirai,Y., Hirai,H., Imai,H. and Takahata,N.  
Submitted (19-SEP-2002) Mineyo Iwase, Graduate University for  
Advanced Studies (Sokenai), Department of Biosystems Science;  
Shonan Kokuasamura, Hayama, Kanagawa 240-0193, Japan  
(E-mail:iwase@koryu.wol.soken.ac.jp, Tel:81-468-58-1571,  
Fax:81-468-58-1544)

FEATURES  
source  
1..5712  
Location/Qualifiers  
/organism="Equus caballus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9796"  
/sex="male"  
174..5712  
/gene="AMELX"  
join(174..233,1578..1589)  
/gene="AMELX"  
join(1590..1643,3566..3613,4993..5037,5305..55712)  
/gene="AMELX"  
/product="amelogenin"  
/codon\_start=1  
/protein\_id="BAC6112.1"  
/db\_xref="GI:29126039"



RESULT 3  
AB091791 7425 bp DNA linear MAM 02-MAY-2003  
LOCUS  
DEFINITION  
AB091791 Sus scrofa AMELX gene for amelogenin, partial cds.  
ACCESSION  
AB091791.1 GI:29126034  
KEYWORDS  
Sus scrofa (pig)  
SOURCE  
Sus scrofa  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.  
REFERENCE  
1 Iwase, M., Satta, Y., Hirai, Y., Hirai, H., Imai, H. and Takahata, N.  
From the cover: The amelogenin loci span an ancient pseudautosomal  
boundary in diverse mammalian species  
Proc. Natl. Acad. Sci. U.S.A. 100 (9), 5258-5263 (2003)  
JOURNAL  
MEDLINE  
22608569  
PUBMED  
12672962  
REFERENCE  
2 (bases 1 to 7425)  
Iwase, M., Satta, Y. and Takahata, N.  
Direct Submision  
Submitted (19-SEP-2002) Mineyo Iwase, Graduate University for  
Advanced Studies (Sokendai), Department of Biosystems Science;  
Shonan Kokusaiinura, Hayama, Kanagawa 240-0193, Japan  
(E-mail: iwase@shonan.ac.jp, Tel: 81-468-58-1571,  
Fax: 81-468-58-1544)  
Location/Qualifiers  
1. .7425  
/organism="Sus scrofa"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9823"  
/sex="male"  
1491. .7425  
/gene="AMELX"  
join(1491. .1548,3140. .3151)  
/gene="AMELX"  
join(3152. .3205,5316. .5363,6619. .6660,6717. .6761,  
7033. .7425)  
/gene="AMELX"  
/codon\_start=1  
/product="amelogenin"  
/protein\_id="BAC66110.1"  
/db\_xref="GI:29126034"  
/translation="MGWILIFACLLGAARMPRLPHGHGYNFSEYEDLYEAIRID  
RLAVLPLKQYQNMRIHRYTSGYBGMGMLHNOIIPVSSQTPQSHLOPHNHLPM  
VPAQPGIPQGPMPRLPGHSMPTQHNQNPRLPRAQPFQPGVQPPHQPQPSPP  
MHPIQPLPQPLPMPFMSQSLPDLPLKMPAT"  
ORIGIN  
Query Match 26.4%; Score 181.2; DB 4; Length 7425;  
Best Local Similarity 67.0%; Pred. NO. 1.8e-29;  
Matches 391; Conservative 0; Mismatches 158; Indels 35; Gaps 8;  
QY 1 ATGGGACCGGATTTTGTTCCTGCTGGAGACCTTCAGTATGCTGAGT 60  
DB 3152 ATGGGACCGGATTTTGTTCCTGCTGGAGACCTTCCTATGCTGTAAGT 311  
QY 61 AAAATTTCC-----AATTTCAATTTCAAGCTTGGAATAATAATGCGCT 108  
DB 3212 AAAGCAACCCCTTCGATACGATGTCATGTCATGCTTGGAATAATAATGCAACC 3271  
QY 109 CAGATTTTATTAAGGTTAAATTAAGGTTTAAAGACATGACATGCTGATGTTTACA 168  
DB 3272 CCAGATT-----GATACAGTTAAGGATGAAAAACATGAAGATCAATTTCTCTCA 3322  
QY 169 AATGTGCTGTGTTAAGAAACATTTTAAATCTTGATCAACAAGATTTTAAAAAATAA 228  
DB 3323 AATGACTCACTGTTTAAAGAAACATTTTAAAGATTTTAAAGATTTTAAAGATTTTAA 3374  
QY 229 CATTCAGTTGCTTCAACCAATATTCATTCAGTATGACGAGAGTGGGGGGTGTG-C 286  
DB 3375 GATTCCCAAGATGCTTCCCAAGATTTCTGATTCAGTATGCTGGGGGGTGGGCCCATGAC 3434

QY 287 TGAAGACTCTGAATTTTAAACAGCAAGACCTGAGAAATTTTGAACAAATTACTTAT 346  
DB 3435 TCTGCATTTTAAACAGCACCCCGAGAGACTCTTAAACAAATTACTTGAATATCAT 3494  
QY 347 AATCACTCATCTAGATGAGAAATTTTGAAGGACCTTTGAAGGCTCC--AGAA 404  
DB 3495 CACCTGCTCTAGAGGTGAACT--TTTGAAGTACCTTTGAGAGGCTCCAGAGAA 3553  
QY 405 AAGTCTCAATCACTCTTTAAGTACTACAAATATGACATTTTCTTAAACCAATTTTC 464  
DB 3554 TACTGAAGACGCTTTAGCAATATGACAAATATGACATTTTCTTAAATCAATTTTC 3613  
QY 465 TCAAGATATCTACTCTTTCTGCTCAGAAAACCCCGGCTATCATCTTCAAG-T 523  
DB 3614 TCCCAAGTCCCAATATCTTCTGCTTCCCAACCCCGCTGTATTTCAAGAGAG 3673  
QY 524 AAGATTCACTGTTAAGAAATTTCTTCTGCACTATTTATTTGT 567  
DB 3674 TAAAGATTGTTAGAAATTTCTTCTTGAACCATCTCTGT 3717  
RESULT 4  
AB091781 6465 bp DNA linear PRI 02-MAY-2003  
LOCUS  
DEFINITION  
AB091781 Pan troglodytes AMELX gene for amelogenin, partial cds.  
ACCESSION  
AB091781.1 GI:29126015  
KEYWORDS  
SOURCE  
Pan troglodytes (chimpanzee)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Pan.  
REFERENCE  
1 Iwase, M., Satta, Y., Hirai, Y., Hirai, H., Imai, H. and Takahata, N.  
From the cover: The amelogenin loci span an ancient pseudautosomal  
boundary in diverse mammalian species  
Proc. Natl. Acad. Sci. U.S.A. 100 (9), 5258-5263 (2003)  
JOURNAL  
MEDLINE  
22608569  
PUBMED  
12672962  
REFERENCE  
2 (bases 1 to 6465)  
Iwase, M., Satta, Y. and Takahata, N.  
Direct Submision  
Submitted (19-SEP-2002) Mineyo Iwase, Graduate University for  
Advanced Studies (Sokendai), Department of Biosystems Science;  
Shonan Kokusaiinura, Hayama, Kanagawa 240-0193, Japan  
(E-mail: iwase@shonan.ac.jp, Tel: 81-468-58-1571,  
Fax: 81-468-58-1544)  
Location/Qualifiers  
1. .6465  
/organism="Pan troglodytes"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9598"  
/sex="male"  
927. .6465  
/gene="AMELX"  
join(927. .982,2283. .2294)  
/gene="AMELX"  
join(2295. .2348,4284. .4331,5607. .5648,5740. .5781,  
6052. .6465)  
/gene="AMELX"  
/codon\_start=1  
/product="amelogenin"  
/protein\_id="BAC66101.1"  
/db\_xref="GI:29126016"  
/translation="MGWILIFACLLGAARMPRLPHGHGYNFSEYEDLYEAIRID  
RLAVLPLKQYQNMRIHRYTSGYBGMGMLHNOIIPVSSQTPQSHLOPHNHLPM  
VPAQPGIPQGPMPRLPGHSMPTQHNQNPRLPRAQPFQPGVQPPHQPQPSPP  
MHPIQPLPQPLPMPFMSQSLPDLPLKMPSTDKX"  
ORIGIN  
Query Match 23.8%; Score 162; DB 9; Length 6465;  
Best Local Similarity 65.3%; Pred. NO. 2.8e-25;

	Matches	385;	Conservative	0;	Mismatches	155;	Indels	50;	Gaps	8;
Oy	1	ATGGGAGACTGGATTTTGTTCGCTCCCTCGGAGACACCTTCAGATGCTGTGAGT	60							
Db	2295	ATGGGAGACTGGATTTTGTTCGCTCCCTCGGAGACAGCTTTTGCCATGCTGTGAGT	2354							
Oy	61	AAATATTC-----CAATTCCAATTGCAAGCTTGGAAATTAATAATGCT	108							
Db	2355	AAATACACCCCTGCATTAAGTCAGTGTCCAAATTCGACAGCTTGGACATTAATAATGCTC	2413							
Oy	109	CAGATTTTATTTAAGGTTAAATTTAAGGTTTAAAGACAGTACAGATCTGATGTTTACA	168							
Db	2414	-----CATAGTGTGTAATAATTAGGTTTAAACAGATAGATCAGATGTCTTCA	2463							
Oy	169	AATGTCATGTGTTTAAAGAAACATTTTAAATCTTGATCACAGATTTTAAACAAAAA	228							
Db	2464	TATGCTCCGGTTGAAAGAAACACTTCAGAGCTT-----GTTTTTAAAGGTA	2512							
Oy	229	CATTCTCAGTTGCTTCACCCATATTCTGATTCAGTATAGCAGATGGGGGGTGTGTG	288							
Db	2513	TATTCCTAAATGCGCCTACCAAAA-----ATTCTATTGGTACAGCTGGGGCGGGGCC	2567							
Oy	289	AGAACTGAAATTTTAAACAGCA---AGAACTCAGGAAATTCCTGGAACAAATTACTT	344							
Db	2568	AGAACTGCAATTTTAAAGACCCGAGAGATCTGTGGAACTGTAGCTGTGTAAT	2627							
Oy	345	ATATTCATCTATCTCTAGATGA--AAATTTTGAAGAGACCTTGAAGAGCCTTCAGA	402							
Db	2628	ATCACACCCACCTCTATAGATGAAGAACCTTTTGAAGAGACCTTGAAGAGTCTCCAGA	2687							
Oy	403	AAAAGTCTCATCAGTCTTT---AAGTACTACAAAATGACAGTTTCTCTAAACCAA	459							
Db	2688	GAAAGTCTTAAACAAGCTTTGGACAAATATTACAGAGTGCAGTTTGTCTTAAACCCA	2747							
Oy	460	ATTTCACAAAGTATCTACTCTCTTCTGCTCAGAAAAACCCCGGCTATCATTTCTCA	519							
Db	2748	ATTCTCTCAAGATTCCAAATCTCTTCCTCGCCCTCCACATATTGCTGCTCTTATCCCTCA	2807							
Oy	520	--GAGTAGATTTCAGTGTAGGAATTCCTCTCCGACATATTATTTGT	567							
Db	2808	GGGGTAAAGATTTTGTGTAGAAATCACTTTTGAACCACTTTCCTGT	2857							

	RESULT 5			
AY040206	LOCUS	AY040206		
	DEFINITION	AY040206 Homo sapiens amelogenin precursor (AMELX) gene, complete cds.	8810 bp	DNA linear PRI 10-APR-2002
	ACCESSION	AY040206		
	VERSION	AY040206.1	GI:15028582	
	KEYWORDS	.		
SOURCE		Homo sapiens (human)		
ORGANISM		Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 8810) Hart,P.S., Hart,T.C., Simmer,J.P. and Wright,J.T. A nomenclature for X-linked amelogenesis imperfecta <i>Arch. Oral Biol.</i> 47 (4), 255-260 (2002) 21920287 MEDLINE PUBMED 11922868		
REFERENCE		2 (bases 1 to 8810) Hart,S., Hart,T.C. and Wright,T. Direct Submission Submitted (14-JUN-2001) Human Genetics, University of Pittsburgh, 3550 Terrace St., 572A Scaife Hall, Pittsburgh, PA 15090, USA Location/Qualifiers 1..8810		
AUTHORS	TITLE	JOURNAL		
FEATURES	SOURCE			

gene

```
/organism="Homo sapiens"  
/mol_type="genomic DNA"  
/db_xref="taxon:9606"  
/chromosome="X"  
/map="Xp22.3-p22.1"  
1..8791
```

promoter	/gene="AMELX" 1..1416 /gene="AMELX"
TATA_signal	1386..1391 /gene="AMELX" join(1417..1472,2781,.2846,4783,.4830,6240,.6281, 6552,.6977,8611,.8791) /gene="AMELX"
mRNA	/product="amelogenin precursor" 1417..1472 /gene="AMELX"
exon	/number=1 2781..2846 /gene="AMELX"
CDS	/number=2 join(2793..2846,4783,.4830,6240,.6281,6552,.6977, 8611,.8616) /gene="AMELX"
exon	/codon_start=1 /product="amelogenin precursor" /protein_id="AAK77213.1" /db_xref="gi:15028583" /translation="MGTWILFACULGAFAAMPRLPRHPCGYINFSYEVLTPLKNVQD IRPYPSSGYPEGMGLNHQIIPVLSDQHPTNTLGQRNNIIIVVDAGQRFVLRQDMMI VPDGWSTPIQHNOBPNLPRAQORQRPVRQNRNQRRGRVNHRMGRRLRQRRLRPL MFNPQLRPMLPDLTLEMPSTDKKEVD"
sig_peptide	2793..2840 /gene="AMELX"
exon	4783..4830 /gene="AMELX"
exon	/number=3 6107..6148 /gene="AMELX"
exon	/number=4 6240..6281 /gene="AMELX"
exon	/number=5 6552..6977 /gene="AMELX"
exon	/number=6 8611..8791 /gene="AMELX"
polyA_signal	/number=7 8785..8791 /gene="AMELX"
ORIGIN	

Query Match	23.6%;	Score 160.4;	DB 9;	Length 8810;
Best Local Similarity	65.1%;	Pred. No. 6e-25;		
Matches	384;	Conservative	0;	Mismatches 156; Indels 50; Gaps 8;

  

QY	1	ATGGGAGACTGGATTTTGTGTCCTGCCTCGTGGAGACGCTTCAGTAGCCTGTGAGT	60
Db	2793	ATGGGAGACTGGATTTTATTTGCTGCCTCGTGGAGACGCTTTGTCATCCTGTGAGT	2855
QY	61	AAATTTTC-----CAATTCCAAATTTCACAAGCTTGGAAATTAATAATCTGCCT	108
Db	2853	AAAAACCCCTGCATATGTCAGTGTCCAAATTTTCACAACTTGGACATTAATAATCTGC--	29110
QY	109	CAGATTTTATTATTTAGGGTTAAATTTAAAGGTTTAAAGACAGTACAAAGATCTGATGTTCA	168
Db	2911	-----TCATGTTGTGTAATTTAGGGTTTAAACAGTATGAGATCAATGTCCTTCA	2961
QY	169	AATGTGACTGTGTTTAGAAACATTTTAAATCTTGATCACAAGATTTTAAACAAAA	228
Db	2962	TATGTCTCTGGGTGAAAGAAACACTTCAGAGCTT-----GTTTTAAAGAGTA	3010
QY	229	CATTCTCAGTTGCCCTTACCCATATTTCTGATTCAGTATAGCCAGAGTGGGGGTGTGCTG	288
Db	3011	TATTTCTCAATATGCCGTACCAAAA-----ATTCTGATTTTGGTATACAGCTGGGGCGGGGCC	3065
QY	289	AGAACTCTGAATTTTAAACAGCA---AGAACTCAGAGAAATCTTGGAAACAAATTACTT	344

Db 3066 AGACCTGCTGATTTTAAAGACCCAGAGATTCGTGGAACTGTAGCTGTAAAT 3125  
Qy 345 ATATCTCTCATCTCTTGATGGA--AAATTTTGAAGGACCTTTGAAAGGCTCCAGA 402  
Db 3126 ATCCACCCATCTCTGATGAGAGAGCTTTTGAGAGGACCTTTGAAAGGTCTCAGA 3185  
Qy 403 AAAAGTCTCATCTCTT---AGTACTACAAAATGCAAGTTTCTCTTAAACCA 459  
Db 3186 GAAAGTCTTACACGCTTTGACAAATTTTACAGAGATTCGAGTTTGTCTTAAACCA 3245  
Qy 460 ATTCTGCAAGATATCTACTCTCTCTGCTCAGAAACCCGCGCTATCATCTTCA 519  
Db 3246 ATTCCTCTCAAGATTCAAATCTCTCTGCTCCGCCACATATGCTGCTCTACCCCTCA 3305  
Qy 520 --GAGTAAAGTTTCAAGTTTGAAGATTTCTCTGCTGACATTTATTTGT 567  
Db 3306 GGGGGTAAAGATTTTGTGTAGAAATTCACATTTTGAACCAATTCCTGT 3355

RESULT 6  
AL357130/c 158142 bp DNA linear HTG 10-JUN-2001  
LOCUS Homo sapiens chromosome X clone RP13-169E15, 4 unordered pieces.  
DEFINITION AL357130.3 GI:9863814  
ACCESSION  
VERSION HTG; HTGS\_PHASE1; HTGS\_CANCELLED.  
KEYWORDS Homo sapiens (human)  
SOURCE  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1  
AUTHORS McLay, K.  
TITLE Direct Submission  
JOURNAL Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
COMMENT On Aug 21, 2000 this sequence version replaced gi:9214076.  
----- Genome Center  
Center: Sanger Centre  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: humquery@sanger.ac.uk  
----- Project Information  
Center project name: B169E15  
----- Summary Statistics  
Assembly program: XGAP4; version 4.5  
Sequencing vector: plasmid; L08752; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Consensus quality: 155448 bases at least Q40  
Consensus quality: 157038 bases at least Q20  
Insert size: 157842; sum-of-contigs  
Insert size: 160705; 33.4% error; agarose-fp  
Quality coverage: 4.56x in Q20 bases; sum-of-contigs Quality coverage: 4.56x in Q20 bases; agarose-fp  
-----  
\* NOTE: This is a 'working draft' sequence. It currently consists of 4 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.  
1 68897: contig of 68897 bp in length  
\* 68898 68997: gap of 100 bp  
\* 68998 122842: contig of 53845 bp in length  
\* 122843 123942: gap of 100 bp  
\* 123943 125584: contig of 6642 bp in length  
\* 125585 129685: gap of 100 bp  
\* 129686 158142: contig of 28456 bp in length.  
FEATURES  
Source 1. 158142  
Location/Qualifiers

/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/Chromosome="X"  
/clone="RP13-169E15"  
/clone\_id="RP13-169E15"  
1. 68897  
/note="assembly; fragment:01108  
fragment\_chain:1  
clone\_end:17  
vector\_side:left"  
misc\_feature 68998..122842  
/note="assembly; fragment:00101  
fragment\_chain:1"  
misc\_feature 122943..129584  
/note="assembly; fragment:00361  
fragment\_chain:1"  
misc\_feature 129685..158142  
/note="assembly; fragment:01285  
fragment\_chain:1  
clone\_end:56  
vector\_side:right"  
ORIGIN  
Query Match 23.6%; Score 160.4; DB 2; Length 158142;  
Best Local Similarity 65.1%; Pred. No. 4e-25;  
Matches 384; Conservative 0; Mismatches 156; Indels 50; Gaps 8;  
Qy 1 ATGGGAGCTGATTTTGTGCTGCTGCTGAGAGAGCTTCAATGATCTGTGAGT 60  
Db 44325 ATGGGAGCTGATTTTGTGCTGCTGCTGAGAGAGCTTTCGATGCTGTGAGT 44266  
Qy 61 AAAATTTT-----CAATTCGAATTTTCAAGCTTGAATTAATTTTCTGCT 108  
Db 44265 AAAAACCCTTTCATAGCAAGTGCATTTTCAAACTTGACATTAATAATTCG-- 44208  
Qy 109 CAAATTTTATTAGGTTAAATTTAAGGTTTAAAGCTTAAAGCTTAAAGCTTGA 168  
Db 44207 -----TCATAGTTGTGTAATTTTAAAGCTTAAAGCTTAAAGCTTGA 44157  
Qy 169 AATGACCTGCTGTTTAAAGAAACATTTTAAATCTTGATCAAGAATTTTAAACA 228  
Db 44156 TATGCTCTGCTGTTTAAAGAAACCTTCAAGAGCTT-----GTTTAAAGGTA 44108  
Qy 229 CATTCGATGCTGCTTCAACCATATTTCTGATTCAGTATTCAGAGTGGGGTGTG 288  
Db 44107 TATTCGAAATTCGCTTCAACCAAA-----ATTCGATTTGTTACAGTGGGGG 44053  
Qy 289 AGAATCTGAATTTTAAAGCA-----AGAACTCAGAAATCTTGAAACAATTA 344  
Db 44052 AGAATCTGAATTTTAAAGCA-----AGAACTCAGAAATCTTGAAACAATTA 43993  
Qy 345 ATATCTCATCTCTAGATGGA--AAATTTTGAAGGACCTTTGAAAGGCTCCAGA 402  
Db 43992 ATACCAACCATCTCTAGATGGAAGCTTTTGAAGGACCTTTGAAAGGTCTC 43933  
Qy 403 AAAAGTCTCATCTCTT---AGTACTACAAAATGCAAGTTTCTCTTAAACCA 459  
Db 43932 GAAAGTCTTAAACGCTTTGACAAATTTTACAGAGATTCAGTTTGTCTTAAAC 43873  
Qy 460 ATTCTGCAAGATATCTACTCTCTCTGCTCAGAAACCCGCGCTATCATCTTCA 519  
Db 43872 ATTCCTCTCAAGATTCAAATCTCTCTGCTCCGCCACATATGCTGCTCTTAC 43813  
Qy 520 --GAGTAAAGTTTCAAGTTTGAAGATTTCTCTGCTGACATTTATTTGT 567  
Db 43812 GGGGGTAAAGATTTTGTGTAGAAATTCACATTTTGAACCAATTCCTGT 43763

RESULT 7  
AC002366 259202 bp DNA linear PRI 11-JUN-1998  
LOCUS Human Xp22 BAC CT-285115 (from Caltech/Research Genetics), PAC  
DEFINITION



RPTC11-27C22 (from Roswell Park Cancer Center), and Cosmid U35B5  
 (from Lawrence Livermore), complete sequence.  
 AC002366 U79549 U70036  
 AC002366.1 GI:2739349  
 HTG.  
 Homo sapiens (human)  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE  
 AUTHORS  
 1 (bases 1 to 259202)  
 Munty,D., Ansari-Lari,M.A., Timms,K.M., Yu,W., Dugan,S., Lu,J.,  
 Shen,Y., Rowland,K., Liu,W., Perez,Y., Haywood,M.,  
 Jain,B., Leal,B., Logan,O., Nguyen,V., Savage,L., Shen,H.,  
 Morley,K., Chen,E., Forcum,J., Aronson,A.D., Chiu,M.W.,  
 Gorrell,J.H., Brundage,E., Di,W., Chinault,C., Nelson,D. and  
 Gibbs,R.A.  
 TITLE  
 JOURNAL  
 Unpublished  
 REFERENCE  
 AUTHORS  
 2 (bases 1 to 259202)  
 Chiu,M.W.  
 TITLE  
 JOURNAL  
 Direct Submission  
 Submitted (23-JUN-1997) Molecular and Human Genetics, Baylor  
 College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 259202)  
 Chiu,M.W.  
 TITLE  
 JOURNAL  
 Direct Submission  
 Submitted (02-JAN-1998) Molecular and Human Genetics, Baylor  
 College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 4 (bases 1 to 259202)  
 Chiu,M.W.  
 TITLE  
 JOURNAL  
 Direct Submission  
 Submitted (11-JUN-1998) Molecular and Human Genetics, Baylor  
 College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 On Jan 2, 1998 this sequence version replaced gi:2642176.  
 Sequencing is completed to a minimum standard of double strand  
 coverage with a minimum of 2 clones and 2 reads with no ambiguities  
 or 2 chemistries with a minimum of 2 clones and 3 reads with no  
 ambiguities. If the sequence quality does not meet this standard,  
 it will be indicated in the annotation.  
 The repeat regions shown were identified using RepeatMasker by  
 Adrian Smit.  
 Sequence similarities were identified using Powerblast by Jinghui  
 Zhang.  
 Exon/Intron boundaries of identified genes were chosen if there  
 were canonical splice junctions that maintained sequence continuity  
 across the splice junctions.  
 FEATURES  
 SOURCE  
 1. 259202  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="X"  
 /map="Xp22"  
 /clone="CT-285115, RPTC11-27C22, U35B5"  
 /complement(1..147)  
 /note="overlap 1..147 in AC004467"  
 misc\_feature  
 repeat\_region  
 /rpt\_family="MIR"  
 complement(1087..1502)  
 /rpt\_family="L1"  
 complement(5292..5579)  
 /rpt\_family="Alu"  
 repeat\_region  
 5302..5553  
 /rpt\_family="SVA"  
 complement(6009..6063)  
 /rpt\_family="L1MC2"  
 complement(7092..7382)  
 /rpt\_family="L1MA5"  
 7423..7721  
 /rpt\_family="Alu"  
 repeat\_region

repeat\_region 9771..10072  
 /rpt\_family="L1MC2"  
 complement(10072..11146)  
 /rpt\_family="THR"  
 repeat\_region complement(11156..11528)  
 /rpt\_family="TRH1b"  
 complement(12944..13151)  
 /rpt\_family="MER20"  
 13862..14010  
 /rpt\_family="MLT1f"  
 14063..14352  
 /rpt\_family="MER42a"  
 15847..16170  
 /rpt\_family="Alu"  
 complement(18743..19117)  
 /rpt\_family="MER7"  
 23507..23614  
 /rpt\_family="L1"  
 23615..24425  
 /rpt\_family="L1PAL1"  
 complement(25475..25968)  
 /rpt\_family="MLT1d"  
 27234..27683  
 /note="similar to EST with GenBank Accession Number  
 H48827"  
 /db\_xref="dbSTS:G28372"  
 complement(28186..28366)  
 /rpt\_family="MER20"  
 28715..28995  
 /rpt\_family="Alu"  
 complement(30217..30408)  
 /rpt\_family="L1MA5"  
 31195..31364  
 /rpt\_family="Alu"  
 complement(33939..34460)  
 /rpt\_family="L1MA2"  
 34952..35384  
 /rpt\_family="MLT1d"  
 36035..36126  
 /rpt\_family="MIR2"  
 36850..37035  
 /rpt\_family="MER20"  
 38042..38250  
 /db\_xref="dbSTS:L47508"  
 38122..38510  
 /rpt\_family="MSTa"  
 38797..40103  
 /rpt\_family="MIR"  
 complement(42465..42539)  
 /rpt\_family="MIR"  
 47135..47335  
 /rpt\_family="MER3"  
 complement(48933..49166)  
 /rpt\_family="MIR"  
 complement(49718..49918)  
 /rpt\_family="L1PAL1"  
 complement(51943..52505)  
 /rpt\_family="L1MB3"  
 53226..53677  
 /rpt\_family="MLT1d"  
 complement(54846..54966)  
 /rpt\_family="MIR"  
 55040..55226  
 /rpt\_family="MER20"  
 55691..55837  
 /rpt\_family="Alu"  
 56051..56486  
 /rpt\_family="L1ME3a"  
 58557..59030  
 /rpt\_family="MLT1d"  
 complement(59270..59428)  
 /rpt\_family="MER5"  
 complement(59847..60030)  
 repeat\_region



```

repeat_region      /rpt_family="MER20"
                    complement(61290..61384)
                    /rpt_family="L1PB1"
repeat_region      61683..61811
                    /rpt_family="MER20"
repeat_region      61907..62113
                    /rpt_family="L1MB7"
repeat_region      complement(63588..63969)
                    /rpt_family="L1MA10"
repeat_region      64800..65073
                    /rpt_family="Alu"
repeat_region      complement(67157..67209)
                    /rpt_family="MIR"
gene               67637..75845
                    /gene="AMELX"
promoter           67637..68496
                    /gene="AMELX"
                    /note="similar to bovine amelogenin promoter"
mRNA               join(68495..68552,69861..69926,71063..71910,73320..73361,
                    73632..74057,75691..75845)
                    /gene="AMELX"
                    /product="amelogenin"
                    /join(69873..69926,71863..71910,73320..73361,73632..74057,
                    75691..75696)
                    /note="similar to amelogenin encoded by GenBank Accession
                    Number M86932"
                    /codon_start=1
                    /product="amelogenin"
                    /protein_id="AAC21581.1"
                    /db_xref="GI:2739350"
                    /translation="MGTWILFACLGAFAMPPLPRPHGPIYNSYEVLTPLKWTOS
                    IRPVPSTYGEPMGMLHQITLVLSQDHPRTTLQPHNIIIVPVAQDPVTPQPMWP
                    VEQHSTNPIQHQPMLPRPAQOPVOPORPQPHQPMOPQPVNHPMQLPPLP
                    MPMQPLPPLPDLTLEAMPSTDKTKKEEVD"
repeat_region      complement(77161..77331)
                    /rpt_family="MER3"
repeat_region      complement(79184..79473)
                    /rpt_family="Alu"
repeat_region      complement(79881..80169)
                    /rpt_family="Alu"
repeat_region      80196..80253
                    /rpt_family="MER20"
repeat_region      80384..80459
                    /rpt_family="L1PB3"
repeat_region      complement(81544..81633)
                    /rpt_family="MIR"
repeat_region      84789..84957
                    /rpt_family="L1MA2"
repeat_region      84978..85252
                    /rpt_family="Alu"
repeat_region      85294..85692

Query Match      23.6%; Score 160.4; DB 9; Length 259202;
Best Local Similarity 65.1%; Pred. NO. 3.7e-25;
Matches 384; Conservative 0; Mismatches 156; Indels 50; Gaps 8;

```

```

QY 229 CATTCAGTGGCTTACCCCAATTCGATTCAATATGCGAGTGGGGGTGCTG 288
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 70091 TATCTCAATGCGCTACCAAAA-----ATTCTGATTTGGTCACTGGGCGGGCCC 70145
QY 289 AGAAGCTGAAATTTTAAACAGCA---AGAACTCAGAAATCTTGAAACAAATTA 344
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 70146 AGAGCTGCAATTTTATAGAGCAAGGAGAGTGTGTTGGAAGCTTGAATTAAT 70205
QY 345 ATATACATCATCTCTAGATGCA--AAATTTTGAAGGAGCTTTGAAAGCCTCAGA 402
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 70206 ATCAACCAATCTCTAGATGAGGAGAGCTTTTGAAGGAGCCTTGAAAGCTCAGA 70265
QY 403 AAAAGTCTCAATCACTCTT---AGTACTCAAAAATGCGAGTTTCTTAAACCA 459
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 70266 GAAAGTCTTACCACTTTGACAAATATTAACAGAGAGCCAGTTTGTCTTAAACCA 70325
QY 460 ATTTCTACAAGATATCACTCTCTTCTGCTCAGAAAACCCCGGCTATCAATTTCA 519
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 70326 ATTCCTCAAGATTCGAATCTTCTGCTCCTCCACATATGCTGCTTACCCCTCA 70385
QY 520 --GAGTAAGATTCAGCTTTAGGAATTTCTCTCTGAACTATTATTGT 567
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 70386 GGGGTAAGATTTTGTGTAGGAATCACTTTTGAGCCACATCTCTGT 70435

```

```

RESULT 8
AB091785 5684 bp DNA linear PRI 02-MAY-2003
LOCUS Lemur catla AMELX gene for amelogenin, partial cds.
DEFINITION AB091785
ACCESSION AB091785
VERSION AB091785.1 GI:29126023
KEYWORDS
SOURCE
ORGANISM

```

```

REFERENCE
AUTHORS Iwase,M., Saita,Y., Hirai,Y., Hirai,H., Imai,H. and Takahata,N.
TITLE Boundary in diverse mammalian species
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 100 (9), 5258-5263 (2003)
MEDLINE 22608569
PUBMED 12672962
REFERENCES
2 (bases 1 to 5684)
AUTHORS Iwase,M., Saita,Y. and Takahata,N.
TITLE Direct Substiation
JOURNAL Submitted (19-SEP-2002) Mineyo Iwase, Graduate University for
Advanced Studies(Soken dai), Department of Biosystems Science;
Shonan Kokusaijima, Hayama, Kanagawa 240-0193, Japan
(E-mail:iwase@koryu.wol.soken.ac.jp, Tel:81-468-58-1571,
Fax:81-468-58-1544)

```

```

FEATURES
source
Location/Qualifiers
1..5684
/organism="Lemur catla"
/mol_type="genomic DNA"
/db_xref="taxon:9447"
/sex="male"
100..5684
/gene="AMELX"

```

```

gene
100..5684
/gene="AMELX"
join(100..156,1503..1514)
/gene="AMELX"
5'UTR
join(1515..1566,3465..3512,4804..4845,4937..4981,
5229..5684)
/gene="AMELX"

```

```

CDS
/codon_start=1
/product="amelogenin"
/translation="MGTWILFACLGAFAMPPLPRPHGPIYNSYENSHGQALISY
GTELVTPLKWTOSMIRPVPSTYGEPMGMLHQITLVLSQDHPRTTLQPHNIIIV
VOAQDPVTPQPMWPVEQHSTNPIQHQPMLPRPAQOPVOPORPQPHQPMOPQPV
NHPMQLPPLPDLTLEAMPSTDKTKKEEVD"

```

ORIGIN

Query Match 21.4%; Score 145.2; DB 9; Length 5684;  
 Best Local Similarity 63.4%; Pred. No. 1.3e-21;  
 Matches 374; Conservative 0; Mismatches 168; Indels 48; Gaps 8;

Db 1 ATGGGAGCTGGATTTTGTGCTGCTCTCGGAGAGAGCTTCAGATGCTGTAGT 60  
 1515 ATGGGAGCTGGATTTTGTGCTGCTCTCGGAGAGAGCTTCAGATGCTGTAGT 1574  
 Qy 61 AAAATTTT-----CAATTCGAATTTCAAGAGCTTGAATAAATCTGCT 108  
 1575 AAAACACCCCTGCAATAGTGTGCAATTTCAAGAGCTTGAATAAATCTGCT 1634  
 Db 109 CAGATTTTATTTAGGGTTAAATTTAAGGGTTAAGCAGTCAAGATCTGATGTC 168  
 1635 CACCGCT-----GGTAACTTTAGGGTTAAATGATATATCAGATGCTCTCA 1685  
 Qy 169 AATGTGACTGTGTTAAGAAACATTTTAAATCTGTATCAGATTTTAAACAAAA 228  
 1686 AAGCTCTCCGTGTGAAGAAACA-----CTTGAAGAGATCGTTTAAAGATTAAG 1735  
 Db 229 CATTCGACTGCTTCAACCATATTTGATTCAGTATAGCCAGAGTGGGGGCTGCTG 288  
 1736 ATTCGCAATGCTCTGCAAAACATCTGATTTGTGTAAGCTGAAGCAGG-----GCC 1789  
 Qy 289 AGAAGCTGCAATTTTAAACAAGCA-----AGAACCTCAGAAATTTTGAACAATTA 344  
 1790 AGAAGCTGCAATTTTAAACAAGCA-----AGAACCTCAGAAATTTTGAACAATTA 1849  
 Db 345 AATACTACTATCTCTAGATGA--AAATTTTGAAGGAGCCTTTG--AAAGGCTCTCA 400  
 1850 ATCACTGCTGCTCTGATGAGAGAGTCTTGAAGAGGAGCCTTTGAAAGGCTCTT 1909  
 Qy 401 GAAAAATGCTCATAGTCTT---TAAGTACTCAAAAATGCCAGTTTCTTAAACC 457  
 1910 GAGAAATGCTTATAGCATTAGGCAAAATATTAACAACATCAAGTTTCTTAAACC 1969  
 Db 458 AAATTTCTCAAGATATCTAATCTCTTTCTGCTCAGAAAACCCCGGCTATCATCT 517  
 1970 CAATTCCTCAAGTTTCCAAATTTCTTCCAGCCCTCCACATACCTCTGATCTTCA 2029  
 Qy 518 CAGAGTAAAGATTCAGTGTAGGAATTTCTTCTGTAATTAATTTGT 567  
 2030 GGGAGTAAATTTCTGTTAGGAATCCCTGCTGAACATCTCTGT 2079

Db 2030 GGGAGTAAATTTCTGTTAGGAATCCCTGCTGAACATCTCTGT 2079

RESULT 9  
 AB091783 6442 bp DNA linear PRI 02-MAY-2003  
 LOCUS Saimiri sciureus AMELX gene for amelogenin, partial cds.  
 DEFINITION AB091783  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE Saimiri sciureus (common squirrel monkey)  
 ORGANISM Saimiri sciureus  
 Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Cebinae;  
 Saimiri.

REFERENCE  
 AUTHORS Iwase, M., Satta, Y., Hirai, Y., Hirai, H., Imai, H. and Takahata, N.  
 TITLE From the Cover: The amelogenin loci span an ancient pseudautosomal  
 boundary in diverse mammalian species  
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 100 (9), 5258-5263 (2003)  
 MEDLINE 22608569  
 PUBMED 12672962  
 2 (bases 1 to 6442)  
 Iwase, M., Satta, Y. and Takahata, N.  
 Direct Submission  
 Submitted (19-SEP-2002) Mineyo Iwase, Graduate University for  
 Advanced Studies (Soken), Department of Biosystems Science,  
 Shonan Kikusaijima, Hayama, Kanagawa 240-0193, Japan  
 E-mail: iwase@min.kozu.wv01.soken.ac.jp, Tel: 81-468-58-1571,  
 Fax: 81-468-58-1544)

FEATURES  
 source Location/Qualifiers  
 1..6442  
 /organism="Saimiri sciureus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9521"  
 /sex="male"  
 936..6442  
 /gene="AMELX"  
 join(936..591,2298..2309)  
 /gene="AMELX"  
 join(2310..2363,4289..4336,5592..5633,5725..5769,  
 6041..6442)  
 /gene="AMELX"  
 /codon\_start=1  
 /product="amelogenin"  
 /protein\_id="BAC66103.1"  
 /db\_xref="GI:29126020"  
 /translation="MGTTILFACLLGAAPMLPDPHGHGYINSEYNSHOAINID  
 RTGLVTLPLKWTOSMIRPAYPSGYEPMGWIHHOILPVLSOHPHTLLOPHNHPV  
 VPAQDPVYQGMNVPYRQSHMIFPHOHQPNIPPPAQPFPQDTVOPOPHQMPQPP  
 VHPQPLPQPLPPLPMPQPLPMLPDLPLAMPAT"

Query Match 20.5%; Score 139.6; DB 9; Length 6442;  
 Best Local Similarity 63.9%; Pred. No. 2.1e-20;  
 Matches 376; Conservative 0; Mismatches 164; Indels 48; Gaps 9;

Db 1 ATGGGAGCTGGATTTTGTGCTGCTCTCGGAGAGAGCTTCAGATGCTGTAGT 60  
 2310 ATGGGAGCTGGATTTTGTGCTGCTCTCGGAGAGAGCTTCAGATGCTGTAGT 2369  
 Qy 61 AAAATTTT-----CAATTCGAATTTCAAGAGCTTGAATAAATCTGCTCA 110  
 2370 AAAACCCCTTGCATTAAGTCAAGTCCAAATTTCAATTAAGTCAATTAAGTCCCA 2429  
 Db 111 GATTTTATTTAGGGTTAAATTTAAGGGTTAAGCAGTAAAGATCTGATGTTCAAA 170  
 2430 -----TGGTGTGAAGTTAGGGTTTAAACAGTATAGATGTCTTCAA 2479  
 Db 171 TGTGACTGTGTTAAGAAACATTTTAAATCTGTATCAGAAATTTAAACAAAA 230  
 2480 CAATTCCTGAGTTGAAGAAACATCTCAGAGCTT-----GTTTAAAGATTA 2528  
 Qy 231 TTCTCAGTGCCTTACCCATATTTGATTCAGTATAGCCAGAGTGGGGGTGTGTG 290  
 2529 TTCTCAATGCTCTCAGCA-----AGATTCGATTTGTATACGTGGGGGTGGCCAG 2583  
 Db 291 AACTGGAATTTTAAACAAGCA-----AGAACTCAGAAATTTCTGAAACAATTA 346  
 2584 GACTGCAATTTTAAAGTACACAGAGATTTGTTGAACATTAAGTCTGTATGAT 2643  
 Qy 347 AATCACTATCTCTAGATGA--AAATTTTGAAGGAGCCTTTGAAGGCTCCAGAA 404  
 2644 CACCAACCAATCTCTAGATGAAGAGAGCTTTGAAGGAGCCTTTGAAGGCTCTAGAA 2703  
 Db 405 AAGTGTCAATGAGCTT---TAAGTACTCAAAAATGCAAGTCTTCTTAACAACAT 461  
 2704 AAGTGTCAATGAGCTTGGGCAAAATTAACAGAGTCCAGTTTGTCTTAACACC-AT 2762  
 Qy 462 TTCTCAGAGTATCAACTCTTTGCTCAGAAAACCCCGGCTATCATCTTCA-- 519  
 2763 TCTCTCAAGTTTCCAAATATCTTCTGCTCCCAATATGTTATTTGATCCCTCAGG 2822  
 Db 520 GAGTAAAGTTCAGGTTAGGAATTTCTTCTGTAATTAATTTGT 567  
 2823 GGGTAAAGTTCAGGTTAGGAATTCATTTCTTGAAGCATCTTCTGT 2870

RESULT 10  
 AB091794 5591 bp DNA linear MAM 02-MAY-2003  
 LOCUS Equus caballus AMELX gene for amelogenin, partial cds.  
 DEFINITION AB091794  
 ACCESSION

```

VERSION      AB091794.1  GI:29126040
KEYWORDS
SOURCE
ORGANISM      Equus caballus (horse)
               Equus caballus
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
REFERENCE
AUTHORS      1
TITLE        Iwase, M., Saita, Y., Hirai, Y., Hirai, H., Imai, H. and Takahata, N.
               From the cover: The amelogenin loci span an ancient pseudoautosomal
               boundary in diverse mammalian species
JOURNAL      Proc. Natl. Acad. Sci. U.S.A. 100 (9), 5258-5263 (2003)
MEDLINE      22608569
PUBMED       12672962
REFERENCE
TITLE        2 (bases 1 to 5591)
AUTHORS      Iwase, M., Saita, Y. and Takahata, N.
               Direct Submission
               Submitted (19-SEP-2002) Mineyo Iwase, Graduate University for
               Advanced Studies (Sokendai), Department of Biosystems Science;
               Shonan Kokusaiimura, Hayama, Kanagawa 240-0193, Japan
               (E-mail: iwase@nkcoryu.w01.soken.ac.jp, Tel: 81-468-58-1571,
               Fax: 81-468-58-1544)
FEATURES
source       Location/Qualifiers
             1..5591
               /organism="Equus caballus"
               /mol_type="genomic DNA"
               /db_xref="taxon:9796"
               /sex="male"
             130..5591
               /gene="AMELY"
               /join(130..180,1483..1494)
               /gene="AMELY"
               /join(1495..1548,3470..3517,4907..4915,5235..5591)
               /gene="AMELY"
               /codon_start=1
               /product="amelogenin"
               /protein_id="BAC66113.1"
               /db_xref="GI:29126041"
               /translation="MGTWILFACLVGTAIAMP.PHPRGHAYINFSYVLTTPSKYE
               PMSGLHQLIIPVLTQONSNNALQPHHHI.PMVPAQQPVSSQPAIPLRQSHMITSQ
               HHOPNLPVPVQSFQPPHOPIDPPHPIOPLPQAPLPPIPLAASGPHAS"
ORIGIN
Query Match      19.4%; Score 135.4; DB 4; Length 5591;
Best Local Similarity 62.4%; Pred. No. 1.8e-19;
Matches 360; Conservative 0; Mismatches 181; Indels 36; Gaps 8;
QY      1  ATGGGAGCTGATTTTGGCTGCTGCTGCGAGCAGCTTCAGTATGCTGAGT 60
DB      1495 ATGGGAGCTGATTTTGGCTGCTGCTGCGAGCAGCTTCAGTATGCTGAGT 1554
QY      61  AAAATTTCCATTTCCATTTCCATTTCCATTTCCATTTCCATTTCCATTTT 120
DB      1555 ATTAAT--CAGCATCTATTTTCCATTTCCATTTCCATTTCCATTTG 1603
QY      121 TAGGGTTAAATTTAAGGGTTAAGACAGTACAGATCTGATCTTCAAAATGACGTG 180
DB      1604 TGTGGTAAATTTTGGGGTTTAAATCAATCAAGATCATATCTTCAAAATGCTCCAG 1663
QY      181 TTTAAGAAACATTTTAAATCTTGATCAACAATTTTAAACAAACAAACATTTCTCAGTTG 240
DB      1664 TTTAAGATTAATTTGAAGAGCT-----AGTTATTAAGAAATAGATTTCTCGAG 1713
QY      241 CCTCAGCCATATTCGATGATATAGCAGAGTGGGGGGTGTCTGAGAACCTTGAAAT 300
DB      1714 ACTCTGCCA-----AGATTTTATGATGAATACGCTGAGAGAGGAGACCCAGGACTGTGTG 1768
QY      301 TTTAACAAGC-----AAGAAGCTCAGGAATTTCTTGAGAAACAATTAATCAATCAAT 356
DB      1769 TTTAATAAGAGAGAGAGAGATCTGTCCACAACATTAAGCTTTTAAAGCTATGAGCCAT 1828
QY      357 CTCTAGATGGA--AAATTTTGGAGAGAGCTTTGAAAGGCTCCAGAAAAAGTGTCTCA-- 413
DB      1829 CTGCAAGAGAGAGAGAGATTTTGAAGGATCTTAAAGGCTCCAGAAATGTTTAAAG 1888

```

```

QY      414 -ATCAGCTTTAGTACTACAAAATATGCCAGTTTCTTTAAACCAATTTCTCAGAGT 472
DB      1889 CAGCCTTAGACAAACATGACACATCATGCTTCTTACCAAGCCAAATTTCTCAGAGT 1948
QY      473 ATCTAAGCTCTTTTGTGCTCAGAAACCCCGGCTATCATTTCTAGA--GTAGATTT 530
DB      1949 GTTTAATTTTCTCCGACTCTCCAACTCCACGCTGTTTCTTTCAGAGGAGAGAAATT 2008
QY      531 CAGTGTAGGAATTTCTTCTGAGTAACTATTTATTGT 567
DB      2009 CTGTGTAGGAATTTCTTATTTAGACATCTCCCTG 2045

RESULT 11
AB091790
LOCUS      AB091790      6264 bp      DNA      linear      MAM 02-MAY-2003
DEFINITION Bos taurus AMELY gene for amelogenin, partial cds.
ACCESSION  AB091790
VERSION    AB091790.1  GI:29126032
KEYWORDS
SOURCE     Bos taurus (cow)
ORGANISM   Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
            Bovinae; Bos.
REFERENCE
AUTHORS    1
TITLE      Iwase, M., Saita, Y., Hirai, Y., Hirai, H., Imai, H. and Takahata, N.
               From the cover: The amelogenin loci span an ancient pseudoautosomal
               boundary in diverse mammalian species
JOURNAL    Proc. Natl. Acad. Sci. U.S.A. 100 (9), 5258-5263 (2003)
MEDLINE    22608569
PUBMED     12672962
REFERENCE
TITLE      2 (bases 1 to 6264)
AUTHORS    Iwase, M., Saita, Y. and Takahata, N.
               Direct Submission
               Submitted (19-SEP-2002) Mineyo Iwase, Graduate University for
               Advanced Studies (Sokendai), Department of Biosystems Science;
               Shonan Kokusaiimura, Hayama, Kanagawa 240-0193, Japan
               (E-mail: iwase@nkcoryu.w01.soken.ac.jp, Tel: 81-468-58-1571,
               Fax: 81-468-58-1544)
FEATURES
source     Location/Qualifiers
             1..6264
               /organism="Bos taurus"
               /mol_type="genomic DNA"
               /db_xref="taxon:9913"
               /sex="male"
               /gene="AMELY"
               /join(114..169,1585..1596)
               /gene="AMELY"
               /join(1597..1659,3515..3562,5080..5124,5863..56264)
               /gene="AMELY"
               /codon_start=1
               /product="amelogenin"
               /protein_id="BAC66109.1"
               /db_xref="GI:29126033"
               /translation="MGTWILFACLVGTAIAMP.PHPRGHAYINFSYVLTTPSKY
               YONMLRYPYSYGYEPVGLHQLIIPVVSQSPQWALQPHHHI.PMVPAQQPVSSQPAIPLRQSHMITSQ
               PMVPAQQPVSSQPAIPLRQSHMITSQ
               PLPIFPMPQPLPVLBDLPLEAMPAT"
ORIGIN
Query Match      19.1%; Score 129.8; DB 4; Length 6264;
Best Local Similarity 62.7%; Pred. No. 3e-18;
Matches 353; Conservative 0; Mismatches 167; Indels 43; Gaps 8;
QY      1  ATGGGAGCTGATTTTGGCTGCTGCTGCGAGCAGCTTCAGTATGCTGAGT 60
DB      1597 ATGGGAGCTGATTTTGGCTGCTGCTGCGAGCAGCTTCAGTATGCTGAGT 1656
QY      61  AAAATTTCCATTTCCATTTCCATTTCCATTTCCATTTCCATTTT 120

```

Db 1657 ATAA--AGCATGACGTATTTTCCAGTTTGAAAAAATCTGCCACAA----- 1706

QY 121 TAGGGTTAAATTTAAGGTTTAAAGACATCAAGATCTGATGTTCAAAATGACTGTG 180

Db 1707 -ATAGGTTAACTTCAGGTTT-----AGAAATTAAGATCAAAATGCTCTAAATGCTCTGTG 1760

QY 181 TTTAAGAAATATTTTAAATCTTGATCACAAGATTTTAAACAAAAAATCTCAGTTG 240

Db 1761 TTTAAGATACAAATTTGAAGAGCTTTGTTATTAATA-----AAAAAAGATCCCAATAT 1816

QY 241 CCTTACCCATATTTCTGATTCAGTATAGCCAGATGGGGGTGTGTAAGACTGTGAT 300

Db 1817 TTCTGTTAAGATTTCTGATTCAGTATAGCTGGGTAG-----TCTAT 1858

QY 301 TTTAACAAGCAAGAACTCAGAAATTTCTGAAACATTTACTTATATCATCTATCT 360

Db 1859 TTTAACAATGAGAGATTCATGATTAACATTTATCTTGAATATATATATACCATCTCT 1918

QY 361 AGATGGA--AATTTTGAAGGGACCTTTGAAGGCTCCAGAAAAAGTCTCAATCAGT 419

Db 1919 AATGAAAGAAATTTTGAAGGACCTCTTCAAGGCTGAAGATTAATTTTATATAT 1978

QY 420 CTTAAGTACTACAAAAATG--CCAGTTTCTTAAACCAAAATTTCTCAAGATATCA 477

Db 1979 GTTAGTAAATACAAATATATAGTTTACCTAAATCAATTTCTTCAAGTTATA 2038

QY 478 ACTCTCTTTCTGCTCAGAAAAACCCGGGCTATCTTCT--AGATTAAGATTTCTG 535

Db 2039 AATATTTTCTGCTCCCTCCAACTCCACAGTCTTATCATCAGAGATTTGCTTCTG 2098

QY 536 TTAGGAATTTCTTCTCTGAACCTA 558

Db 2099 TTTAAATTTCTCTTCTGAAACA 2121

## RESULT 12

AB091787

LOCUS AB091787 5562 bp DNA linear PRI 02-MAY-2003

DEFINITION Ocolemur garnettii AMELX gene for amelogenin, partial cds.

AB091787.1 GI:29126027

VERSION AB091787.1 GI:29126027

KEYWORDS

SOURCE

ORGANISM

Ocolemur garnettii (small-eared galago)

Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Strepsirrhini; Galagonidae; Ocolemur.

1 Iwase, M., Satta, Y., Hirai, Y., Hirai, H., Imai, H. and Takahata, N.

From the Cover: The amelogenin loci span an ancient pseudautosomal

boundary in diverse mammalian species

Proc. Natl. Acad. Sci. U.S.A. 100 (9), 5258-5263 (2003)

JOURNAL MEDLINE 22608569

PUBMED 12672962

2 (bases 1 to 5562)

Iwase, M., Satta, Y. and Takahata, N.

Direct Submission

Submitted (19-SEP-2002) Mineyo Iwase, Graduate University for

Advanced Studies (Sokendai), Department of Bioregions Science;

Shonan Kokusaiinura, Hayama, Kanagawa 240-0193, Japan

(E-mail: iwase@shonan.ac.jp, Tel: 81-468-58-1571,

Fax: 81-468-58-1544)

FEATURES

Location/Qualifiers

1..5562

/organism="Ocolemur garnettii"

/mol\_type="genomic DNA"

/db\_xref="taxon:30611"

/sex="male"

267..5562

/gene="AMELX"

/join(267..321,1401..1412)

/gene="AMELX"

join(1413..1466,3385..3432,4699..4740,4832..4876,

5125..55562)

## ORIGIN

Query Match 18.2%; Score 123.8; DB 9; Length 5562;

Best Local Similarity 60.5%; Pred. No. 6.1e-17;

Matches 389; Conservative 0; Mismatches 197; Indels 57; Gaps 9;

/gene="AMELX"

//codon\_start=1

/product="amelogenin"

/protein\_id="BAC66107.1"

/db\_xref="GI:29126028"

/translation="MGTWLFACLGAAFAWPLPHPGHYINFSYKSYSAQAINID

RAVALVLPVKYQSMILRPYPSYSGYKPMQYHNOIIPVLSQOHPPTLQPHNIPV

VPADQPVDPQPMPLPPOHSMTPPOHSMTPPOHSMTPPOHSMTPPOHSMTPPOH

MPMPQIPQPVHMPPLPPOHSMTPPOHSMTPPOHSMTPPOHSMTPPOHSMTPPOH

QY 1 ATGGGACCTGGATTTTGTGCTGCTGCTGCGAGCGAGCTTCAATATGCTGAGT 60

Db 1413 ATGGGACCTGGATTTTGTGCTGCTGCTGCGAGCGAGCTTCCATCCCGGAGT 1472

QY 61 AAAATTTT-----CAATTCGAATTTCAAGCTTGAATTAATAATCTGCGT 108

Db 1473 AAAACACCCCTTGCAAAAGTCAGTATCAATTTTCAATGATTAATAATAATCTGCTC 1532

QY 109 CAGATTTTATTTAGGTTAAATTTAAGGTTTAAAGACATCAAGATGATTTTCA 168

Db 1533 AACAGTT-----GTAACTTCAGATTTTAAACAGTATATATCAGATGCTCTCA 1583

QY 169 AATGACCTGTGTTTAAAGAAATTTTAAATCTTGATCACAAGATTTTAAACAAAA 228

Db 1584 AACATCTCTCTC-----GTGTGAAGAAACATTTGAGAGATTTACTTAAAGATA 1634

QY 229 CATTCAGTTGCTTCAACCAATATCTGATTCAGTATAGCCAGATGGGGGTGTGCTG 288

Db 1635 GATTTCAAGAGCTCTCACCA-----AGATTCGATTTGTTAAGACAGACAGAGGCC 1689

QY 289 AGAATCTGAATTTTAAACAGC-----AAGAACCTCAGAAATTTCTGAAACATTTACT 343

Db 1690 AGGCTCTGAGTTTCAAGACAGCCTTAGAGATTTCTCATTTGAACATTTAGCTGTAA 1749

QY 344 TATATCATCATCTTATATAGTA--AAATTTTGAAGAGACCTTGAAGAGCTCTCCAG 401

Db 1750 TATCATCTGCCATCTCTATAGTAGAGAGATTTTCAAGAGGCTTCAAGAGCTCTCCAG 1809

QY 402 AAAAGTGTCAATCATGCTTT-----AAGTACTCAAAATATGACGATTTTCTTA 452

Db 1810 AGAAAGTGTATATAGCTTTAGGCAAAATCTTATTAACAAATCAAGTTTCTTA 1869

QY 453 AAACCAATTTTTCACAAGATTTTAACTCTTTTGTCTGAGAAACCC-----CGAGCT 508

Db 1870 AAGCCAGTTTCTTCAAGTTTCAAAATCTCTTCCATCCCTCACAATACATGCTGTT 1929

QY 509 ATATCTTCAAGATTAAGATTTTCAAGATTTTCAAGATTTTCTCTCTGA--ACTATTTTGG 566

Db 1930 ATCTTTTAAAGTGAAGATTTCTGAACCATCTCTTATTTTGAAGGCCCAACCCCTTT 1989

QY 567 TCAGTACCTGCTAAGACATGATCTCTATTTGTTGTTGTCAG 609

Db 1990 GAGTATATCTACTAAGATTTTGGCTTCTTGGTGGTCCAGAG 2032

## RESULT 13

DB3730

LOCUS DB3730 1935 bp DNA linear PRI 10-JUL-2001

DEFINITION Homo sapiens AMGX gene for amelogenin, intron 2.

DB3730.1 GI:5263179

VERSION DB3730.1 GI:5263179

KEYWORDS

SOURCE Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 Sekiguchi, H., Minaguchi, K., Machida, Y. and Yakuhi, M.

TITLE PCR detection of the human amelogenin gene and its application to the diagnosis of amelogenesis imperfecta  
JOURNAL Bull. Tokyo Dent. Coll. 39 (4), 275-285 (1998)  
MEDLINE 99234629  
PUBMED 10218009  
REFERENCE 2 (bases 1 to 1935)  
AUTHORS Sekiguchi, H.  
TITLE Direct Submision  
JOURNAL Submitted (29-FEB-1996) Hiroshi Sekiguchi, Tokyo Dental College, Pediatric Dentistry, 1-2-2 Masago, Mihama-ku, Chiba, Chiba 261, Japan (Tel:043-270-3945, Fax:043-279-2052)  
FEATURES  
source  
1. .1935  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="X"  
1. .1935  
/gene="AMGX"  
1. .1935  
/gene="AMGX"  
/note="AMGX"  
/note="amelogenin"  
this nucleotide sequence shows an intron between exon2 and exon3 within human amelogenin gene on the X chromosome"  
/number=2  
ORIGIN  
Query Match 17.4%; Score 118.4; DB 9; Length 1935;  
Best Local Similarity 63.7%; Pred. No. 1.1e-15;  
Matches 325; Conservative 0; Mismatches 146; Indels 39; Gaps 8;  
QY 69 CAATTCGAATTCAGAGTGGAAATTAATTCGCCAGATTTTATTTAGGTTA 128  
DB 27 CAGGTGCAATTCACAACTTGACATTAATAATTCGCC-----CATAGTTGCT 75  
QY 129 AATTTAGGGTTTAAAGACATGATCATGATTCACAAATGCTGATGTTTAAAGA 188  
DB 76 GAATTAAGGGTTTAAAGACATGATCATGATTCACAAATGCTGATGTTTAAAGA 135  
QY 189 ACATTTTAAATCTTGATCACAAGATTTTAAACAAAACATTCGAGTGGCTTACC 248  
DB 136 ACACCTCAGGAGCTTG-----TTTAAAGATATTTCTCAATGCGCGTACC 184  
QY 249 CATATTCGATTCAGATTAAGCCAGATGGGGGGTGTGCTGAACAATCTGAATTTTAAACA 308  
DB 185 AAAA-----ATCTGATTTGGTACAGCTGGGGGGGCCCGAGACTCTGCATTTTAAATA 239  
QY 309 GCA-----AGAACTCAGAAATCTTGAACAAATTAATTAATCACTCATCTAGAT 364  
DB 240 GCACCCCAGAGATCTGTTGAACTGTGATCTGTAATATCAACACCATCTTCAAGT 299  
QY 365 GGA- -AAATTTTGAAGGAGCCTTGAAGGCTTCAGAAAAGTCTCAATCACTT 422  
DB 300 GGAGGAAGCTTTGGAAGGAGCCTTGAAG- -GTCTCAGAGAAAGTCTTAACCACTTT 358  
QY 423 ----TAAGTACTACAAAATGCGAGTTTCTCTAAACCAATTTCTCAAGATTAAC 479  
DB 359 GGACAAATATTAAGAGATGCGAGTTTGTCTAAACCCCAATTCCTCAAGATTTCAAA 418  
QY 480 TCTCTTTCTGCTCAGAAAACCCCGGCTATCATTTCTTCA- -GAGTAAGATTTCAAGTT 537  
DB 419 TCTCTTCTGCGCTCCACATATGCTGCTTACCCCTCAGGGGGTAAGATTTTGTGTT 478  
QY 538 AGGAATTCCTTCCTGAACCTATTATTGT 567  
DB 479 AGGAATTCACCTTTTGAAGCAATTCCTGT 508  
RESULT 14  
AB091784 7454 bp DNA linear PRI 02-MAY-2003  
LOCUS AB091784  
DEFINITION Saimiri sciureus AMELY gene for amelogenin, partial cds.  
ACCESSION AB091784

VERSION AB091784.1 GI:29126021  
KEYWORDS  
SOURCE Saimiri sciureus (common squirrel monkey)  
ORGANISM Saimiri sciureus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Cebinae; Saimiri.  
REFERENCE  
1 Iwase, M., Saita, Y., Hirai, Y., Hirai, H., Imai, H. and Takahata, N.  
TITLE From the Cover: The amelogenin loci span an ancient pseudautosomal boundary in diverse mammalian species  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 100 (9), 5258-5263 (2003)  
MEDLINE 22608569  
PUBMED 12672962  
REFERENCE 2 (bases 1 to 7454)  
AUTHORS Iwase, M., Saita, Y. and Takahata, N.  
TITLE Direct Submision  
JOURNAL Submitted (19-SEP-2002) Mineyo Iwase, Graduate University for Advanced Studies (Sokendai), Department of Biosystems Science; Shonan Kokusaiimura, Hayama, Kanagawa 240-0193, Japan (E-mail: iwase.mineyo@koryuwo1.boken.ac.jp, Tel: 81-468-58-1571, Fax: 81-468-58-1544)  
FEATURES  
source  
1. .7454  
Location/Qualifiers  
/organism="Saimiri sciureus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9521"  
/sex="male"  
2028. .7454  
/gene="AMELY"  
join(2028. .2083,3075. .3086)  
/gene="AMELY"  
join(3087. .3140,5297. .5344,6741. .6785,7056. .>7454)  
/gene="AMELY"  
/codon\_start=1  
/product="amelogenin"  
/protein\_id="BAC66104.1"  
/db\_xref="GI:29126022"  
/translation="MGTWLIFACLVAAFPAMPLPDPHGGYINFSYEVLTPLKMYOT  
MIRPSYSGYEPMGWMLYHQIIPVLSQHPPTLLOPHHNPVPAOOPVPOOPPM  
PVGQHSMPAHOHPNPPHVOQFPQTVQPOSHQPPQPPVHPVPLPDPPLP  
PMFHPQPLPMLPDLMLKGGQ"  
ORIGIN  
Query Match 14.0%; Score 95.2; DB 9; Length 7454;  
Best Local Similarity 58.7%; Pred. No. 1e-10;  
Matches 223; Conservative 0; Mismatches 148; Indels 9; Gaps 3;  
QY 197 AAAATCTTATCAAGAATTTTAAACAAAACATTCAGTTGCTTCAACCATATTCT 256  
DB 3505 AAAAATTAATAATCAATTAATTAATAATAATAGCTTTTATGATGCTTCACACAGA 3564  
QY 257 GATTCAGTATGACAGAGTGGGGGGTGTGCTGAACCTGAAATTTTAAACAGCAAGAC 316  
DB 3565 GGTTCGATTTGGTACAGTGAAGTTGGGCCATTAATCTGAATTTTAAACAGCACCCAG 3624  
QY 317 CTCAGAAATCTTGAACCAATTAATTAAT- -CACTCATCTAGATGGAATAAT 372  
DB 3625 GAGTTTCTGTGACACCATTAAGCTGTATTAATTAACAACCATTTGATGAAGAAAGCT 3684  
QY 373 TTTGAAGGAGCCTTTGAAGGCTTCGAGAAAGGCTCAATCACTT- -TAAGTAC 429  
DB 3685 TTTGAAGGATCTTTGAAGATCTTCACAGAAAGGCTTAATTAAGCCTTAATAATAATAT 3744  
QY 430 TACAAAATGACAGTTTCTTAACCAATTTCTCAAGATTAATCACTCTTTCTG 489  
DB 3745 AACAAAATACAGATTTGCTTAACAGCAATTCCTCTCAATTTCCAGATCTTTCTTG 3804  
QY 490 CTCAGAAAACCCCGGCTATCATTTCTC- -AGAGTAAGATTTCAAGTGAATTTCC 547  
DB 3805 CCCTCCACATATTATATAGCTTACCTCCTCAGAGAGTAAGATTTCTGTGTTAGAAACCCAC 3864  
QY 548 TTCCTGAACATTAATTTGT 567



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 8, 2005, 11:25:10 ; Search time 902.049 Seconds  
(without alignments)  
3957.220 Million cell updates/sec

Title: US-10-754-437-23

Perfect score: 680  
Sequence: 1 atggggagactgagatttgtt.....ggtactgagagacagacac 680

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 413486 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

N\_Geneseq\_23Sep04:\*  
1: geneseqn1980s:\*  
2: geneseqn1990s:\*  
3: geneseqn2000s:\*  
4: geneseqn2001as:\*  
5: geneseqn2001bs:\*  
6: geneseqn2002as:\*  
7: geneseqn2002bs:\*  
8: geneseqn2003as:\*  
9: geneseqn2003bs:\*  
10: geneseqn2003cs:\*  
11: geneseqn2003ds:\*  
12: geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	160.4	23.6	259202	12	ADQ18492
2	51.2	7.5	722	6	AA141111
3	51.2	7.5	752	6	AA141110
4	49	7.2	270	12	ADM80835
5	49	7.2	349	12	ADM80826
6	49	7.2	623	10	ADM80827
7	45.6	6.7	556	10	ADBS9026
8	45.6	6.7	556	10	ADBS3782
9	45	6.6	476	3	AZS50837
10	44.8	6.6	318	12	ADM80836
11	44.8	6.6	750	2	AAZ07020
12	44.8	6.6	793	2	AAZ07018
13	43.2	6.4	802	2	AAZ07019
14	43.2	6.4	852	12	ADQ22977
15	43	6.3	178896	6	ABO88146
16	41.2	6.1	2000	6	ABZ17518
17	41.2	6.1	15399	6	ABL33515
18	40.6	6.0	2000	8	ADA71938
19	40.4	5.9	110000	5	AA161373
20	39.8	5.9	3001	6	ABK34028
21	39.8	5.9	5118	8	ADA20376

22	39.8	5.9	5118	8	ADA84183	ADA84183 Human ren
23	39.8	5.9	7763	6	ABL34436	ABL34436 Human imm
24	39.8	5.9	110000	6	ABA03041_00	ABA03041 Listeria
25	39.6	5.8	645	4	AA108580	AA108580 Human bre
26	39.6	5.8	3474	4	ABL14636	ABL14636 Drosophila
27	39.6	5.8	7113	6	ABL132805	ABL132805 Human imm
28	39.6	5.8	8979	6	ABL32785	ABL32785 Human imm
29	39.6	5.8	8979	6	ABK31271	ABK31271 Signal tr
30	39.6	5.8	8979	6	ABL70232	ABL70232 Chemical
31	39.6	5.8	8979	6	AA561179	AA561179 Human gen
32	39.6	5.8	80331	4	AA095559	AA095559 Human his
33	39.2	5.8	6282	6	ABL132455	ABL132455 Human che
34	39.2	5.8	6641	6	ABN80003	ABN80003 Human che
35	39	5.7	65	6	ABN28389	ABN28389 Rat eplic
36	39	5.7	2625	4	ABL24966	ABL24966 Drosophila
37	39	5.7	2780	12	AD143260	AD143260 Plant tra
38	39	5.7	5964	6	ABN60170	ABN60170 Human che
39	39	5.7	8456	6	ABL33976	ABL33976 Human imm
40	38.6	5.7	2478	2	AA99552	AA99552 Nucleic a
41	38.6	5.7	8056	8	ABZ10246	ABZ10246 Hematopo
42	38.6	5.7	110000	12	ADQ79173_0	ADQ79173 KLF12 gen
43	38.4	5.6	492	10	ADB81854	ADB81854 Arabidops
44	38.4	5.6	11622	6	ABL32674	ABL32674 Human imm
45	38.4	5.6	16228	6	ABL70459	ABL70459 Chemical

#### ALIGNMENTS

RESULT 1	ADQ18492	ADQ18492 standard; DNA; 259202 BP.
ID	ADQ18492	
XX	AC	ADQ18492;
XX	DT	26-AUG-2004 (first entry)
DE	Human soft tissue sarcoma-upregulated DNA - SEQ ID 1311.	
XX	KM	soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human; de.
OS	Homo sapiens.	
XX	PN	WO2004048938-A2.
XX	PD	10-JUN-2004.
XX	PF	26-NOV-2003; 2003WO-US038193.
XX	PR	26-NOV-2002; 2002US-0429739P.
XX	PA	(PROT-) PROTEIN DESIGN LABS INC.
XX	PI	Aziz N, Ginsburg WM, Zlotnick A;
DR	WPI; 2004-441208/41.	
XX	PT	Early detection of soft tissue sarcoma comprises determining expression of a gene in a first soft tissue sample and a normal soft tissue sample and comparing the gene expression, also useful in treating soft tissue sarcoma.
PT	Example 2; SEQ ID NO 1311; 210pp; English.	
XX	PS	The invention relates to a novel method for detecting soft tissue sarcoma which comprises obtaining a first soft tissue sample from an individual, determining the expression of a gene in both samples and comparing the expression of the gene in both soft tissue samples, where a higher level of protein expression in the first soft tissue sample indicates the presence of soft tissue sarcoma. The method of the invention has cytostatic applications and may be useful for detecting soft tissue



CC sarcoma, possibly via gene therapy or vaccine production. The nucleic  
 CC acid sequences may be useful in diagnostic and screening applications.  
 CC The current sequence is that of a human soft tissue sarcoma up-regulated  
 CC DNA of the invention. The current sequence is not shown within the  
 CC specification per se but was submitted in CD format by the inventor.

XX  
 SQ Sequence 259202 BP; 81699 A; 51421 C; 49221 G; 76861 T; 0 U; 0 Other;

Query Match 23.6%; Score 160.4; DB 12; Length 259202;

Best Local Similarity 65.1%; Pred. No. 3.3e-31; Mismatches 156; Indels 50; Gaps 8;

Matches 384; Conservative 0;

QY 1 ATGGGACCTGGATTTGTTGCTGCTGCTGAGACAGGCTTCAATAGCCCTGAGT 60

DB 69873 ATGGGACCTGGATTTGTTGCTGCTGCTGAGACAGGCTTCAATAGCCCTGAGT 69932

QY 61 AAAATTC-----CAATTCATTTTCAACAAGCTTGGAAATTAATTCGCT 108

DB 69933 AAAACACCCCTTGATAGTCAGTCCAAATTCACAACTTGGACATTAATAATCTGC-- 69990

QY 109 CAGATTTTATTTAGGGTTAAATTAAGGTTTAAAGACAGTCAAGATCTGATGTTGACA 168

DB 69991 -----TCATAGTTGTCGAATTAAGGTTTAAACAGTATGAGATCAGATGCTTCA 70041

QY 169 AATGTGACTGTGTTAAGAAACATTTTAAATCTGATCACAAGATTTTAAACAAAA 228

DB 70042 TATGTCCTGGGTTGAAGAAACACTTCAGAGCTT-----GTTTAAAGGTA 70090

QY 229 CATTTCTAGTTGCTTCAACCATTTTCTGATTCAGTATGACAGAGTGGGGGTGCTG 288

DB 70091 TATTTCTAAATGCGCTACCAAA-----ATTGATTTTGTATCAGCTGGGGGGGGCC 70145

QY 289 AGAAGCTGAATTTTAAACAAGCA---AGAAGCTCAGAAATCTTGGAAACAATTACTT 344

DB 70146 AGAAGCTGAATTTTAAACAAGCAAGAGATCTGTTGGAAGCTTATGCTTGAAT 70205

QY 345 ATAACTACTATCTCTAGATGA--AAATTTTGAAGGAGACCTTGAAGGCTCCGAGA 402

DB 70206 ATCAACACCCATCTCTAGATGAAGGAGACCTTGAAGGAGACCTTGAAGGCTCCGAGA 70265

QY 403 AAAAGTCTAATGAGCTT---AAGTACTACAAAATGCCGTTTCTCTAAACCA 459

DB 70266 GAAAGTCTTACACAGCTTGAACAAATATTAAGAGATGCCAGTTTGTCTAAACCA 70325

QY 460 ATTTCACAAGATTAATCTTCTCTGCTCAGAAAACCCCGGCTATCATTTCTTCA 519

DB 70326 ATTTCCTCAAGATTCAAATCTCTCTGCTCCCAATATGCTGCTTTACCCCTCA 70385

QY 520 --GAGTAAAGATTTCAAGTTAGGAATTCCTTCTGAACTATTTATTTGT 567

DB 70386 GGGGGTAAAGATTTTGTGTAGGAATTCATTTTGAAGCAATTCCTGT 70435

# RESULT 2

AA141111 standard; DNA; 722 BP.

AA141111;

16-OCT-2002 (first entry)

GMML related Y-chromosome DNA sequence.

Goat embryo sexual identification technique; goat amelogenin gene; GMML;

sex-specific; gene; ds; Y-chromosome.

Capra hircus.

Location/Qualifiers

Key 35..658

/\*tag= a

/product= "Y-chromosome protein"

PN TW454013-A.

PD 11-SEP-2001.

XX 10-NOV-1999; 99TW-00119616.

XX 10-NOV-1999; 99TW-00119616.

XX (CHEN/) CHEN C.

XX (JANG/) JANG J.

XX (WENG/) WENG T.

XX (JENG/) JENG D.

PI Chen C, Jang J, Weng T, Jeng D;

DR WPI, 2002-442016/47.

DR P-PSDB; AAO22534.

PT Sex-specific sequence of goat amelogenin gene, useful for embryo sexual

PT identification, comprises high sensitivity even using single white blood

PS cell or cleavage c.

PS Disclosure, Page 28; 35pp; Chinese.

CC The invention relates to a goat embryo sexual identification technique

CC with high efficiency, sensitivity and repeatability. This technique

CC involves separately cloning and sequencing the coding regions and the

CC introns of the goat amelogenin gene (GMML) on the goat chromosomes. The

CC results indicate that there are sex-specific sequences in the fifth

CC intron of the gene. The major characteristics according to the present

CC invention include high sensitivity, applicable in sex identification even

CC only using a single white blood cell or a single cleavage cell of

CC blastula; high diagnostic efficiency, capable of identifying hundreds of

CC goat embryo in 3 hours; simple operation procedures without complicated

CC steps of DNA extraction and need no additional control group intron; and

CC can be applied on different species of goats. This polynucleotide

CC sequence represents a GMML related Y-chromosome DNA sequence of the

CC invention

SQ Sequence 722 BP; 173 A; 264 C; 140 G; 145 T; 0 U; 0 Other;

Query Match 7.5%; Score 51.2; DB 6; Length 722;

Best Local Similarity 94.6%; Pred. No. 0.0025;

Matches 53; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATGGGACCTGGATTTGTTGCTGCTGCTGAGACAGGCTTCAATAGCCCTGT 56

DB 35 ATGGGACCTGGATTTGTTGCTGCTGCTGAGACAGGCTTCAATAGCCCTGT 90

# RESULT 3

AA141110 standard; DNA; 752 BP.

AA141110;

16-OCT-2002 (first entry)

GMML related X-chromosome DNA sequence.

Goat embryo sexual identification technique; goat amelogenin gene; GMML;

sex-specific; gene; ds; X-chromosome.

Capra hircus.

Location/Qualifiers

Key 35..658

/\*tag= a

/product= "X-chromosome protein"



```
XX PF 10-NOV-1999; 99TW-00119616.
XX PR 10-NOV-1999; 99TW-00119616.
XX PA (CHEN/) CHEN C.
XX PA (JANG/) JANG J.
XX PA (WENG/) WENG T.
XX PA (JENG/) JENG D.
XX PI Chen C, Jang J, Weng T, Jeng D;
XX WPI: 2002-442016/47.
XX DR P-PSDB; AAO22534.
XX PT Sex-specific sequence of goat amelogenin gene, useful for embryo sexual
XX PT identification, comprises high sensitivity even using single white blood
XX PT cell or cleavage c.
XX PS Disclosure; Page 28; 35pp; Chinese.
XX CC The invention relates to a goat embryo sexual identification technique
XX CC with high efficiency, sensitivity and repeatability. This technique
XX CC involves separately cloning and sequencing the coding regions and the
XX CC introns of the goat amelogenin gene (GAML) on the goat chromosomes. The
XX CC results indicate that there are sex-specific sequences in the fifth
XX CC intron of the gene. The major characteristics according to the present
XX CC invention include high sensitivity, applicability in sex identification even
XX CC only using a single white blood cell or a single cleavage cell of
XX CC blastula, high diagnostic efficiency, capable of identifying hundreds of
XX CC goat embryo in 3 hours; simple operation procedures without complicated
XX CC steps of DNA extraction and need no additional control group intron; and
XX CC can be applied on different species of goats. This polynucleotide
XX CC sequence represents a GAML related X-chromosome DNA sequence of the
XX CC invention
XX SQ Sequence 752 BP; 184 A; 271 C; 152 G; 145 T; 0 U; 0 Other;
XX
XX Query Match 7.5%; Score 51.2; DB 6; Length 752;
XX Best Local Similarity 94.6%; Pred. No. 0.0026;
XX Matches 53; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 ATGGGGACCTGATTTGTTGCTGCTCGTGGAGACAGCTTCAATATGCCGT 56
DB 35 ATGGGACCTGATTTGTTGCTGCTCGTGGAGACAGCTTCTATGCCCT 90
XX
XX RESULT 4
XX ADM80835
XX ID ADM80835 standard; cDNA; 270 BP.
XX AC ADM80835;
XX DT 03-JUN-2004 (first entry)
XX DE Human CADECM-22 encoding cDNA SEQ ID NO:64.
XX XX human, cell adhesion and extracellular matrix protein; CADECM;
XX KW neuroprotective; cytoskeletal; anorectic; immune disorder;
XX KW neurological disorder; developmental disorder;
XX KW connective tissue disorder; cell proliferative disorder; cancer; obesity;
XX KW Tangle disease; gene; ss.
XX OS Homo sapiens.
XX XX
XX Key Location/Qualifiers
XX FH CDS 36..212
XX FT /*tag= a
XX FT /product= "CADECM-22"
XX PN WO2004015396-A2.
XX PD 19-FEB-2004.
```

```
XX PF 12-AUG-2003; 2003MO-US025418.
XX PR 13-AUG-2002; 2002US-0403781P.
XX PR 30-AUG-2002; 2002US-0407034P.
XX PR 13-SEP-2002; 2002US-0410566P.
XX PR 24-SEP-2002; 2002US-0413482P.
XX PR 25-SEP-2002; 2002US-0413890P.
XX PR 08-NOV-2002; 2002US-0424504P.
XX PR 13-NOV-2002; 2002US-0426222P.
XX PA (INCY-) INCYTE CORP.
XX XX
XX PI Elliott VS, Khare R, Emerling BM, Kable AE, Tran UK, Jin P;
XX PI Becha SD, Margulis JP, Swarnakar A, Chawla NK, Ramkumar J, Blake JJ;
XX PI Hafalia AJA, Lee SY, Jiang X, Jackson AA, Richardson TW, Blake JJ;
XX PI Wang JT, Chien D, Yang YG;
XX WPI: 2004-191795/18.
XX DR P-PSDB; ADM80793.
XX PT New cell adhesion and extracellular matrix proteins, useful in
XX PT diagnosing, treating and preventing immune, neurological, developmental,
XX PT connective tissue and cell proliferative disorders including cancer.
XX PS Claim 5; SEQ ID NO 64; 272pp; English.
XX CC The present sequence encodes a human cell adhesion and extracellular
XX CC matrix protein designated CADECM. CADECM sequences has neuroprotective,
XX CC cytoskeletal and anorectic activities. The CADECM polypeptides and
XX CC polynucleotides are useful in diagnosing, treating and preventing immune,
XX CC neurological, developmental, connective tissue and cell proliferative
XX CC disorders including cancer, e.g. breast, prostate, ovarian, lung or colon
XX CC cancer, obesity and Tangle disease.
XX SQ Sequence 270 BP; 76 A; 65 C; 67 G; 62 T; 0 U; 0 Other;
XX
XX Query Match 7.2%; Score 49; DB 12; Length 270;
XX Best Local Similarity 79.5%; Pred. No. 0.0071;
XX Matches 58; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
QY 1 ATGGGGACCTGATTTGTTGCTGCTCGTGGAGACAGCTTCAATATGCCGTAGT 60
DB 36 ATGGGACCTGATTTATTTATTTGCTGCTCGTGGAGACAGCTTTGCATGCTGCTT 95
QY 61 AAAATTTCCAATT 73
DB 96 ACCCCTTGAAGT 108
XX
XX RESULT 5
XX ADM80826
XX ID ADM80826 standard; cDNA; 549 BP.
XX AC ADM80826;
XX DT 03-JUN-2004 (first entry)
XX DE Human CADECM-13 encoding cDNA SEQ ID NO:55.
XX XX human, cell adhesion and extracellular matrix protein; CADECM;
XX KW neuroprotective; cytoskeletal; anorectic; immune disorder;
XX KW neurological disorder; developmental disorder;
XX KW connective tissue disorder; cell proliferative disorder; cancer; obesity;
XX KW Tangle disease; gene; ss.
XX OS Homo sapiens.
XX XX
XX Key Location/Qualifiers
XX FH CDS 36..491
XX FT /*tag= a
XX FT /product= "CADECM-13"
XX PD 19-FEB-2004.
```

PN WO2004015396-A2.  
 XX  
 PD 19-FEB-2004.  
 XX  
 PF 12-AUG-2003; 2003WO-US025418.  
 XX  
 PR 13-AUG-2002; 2002US-0403781P.  
 PR 30-AUG-2002; 2002US-0407034P.  
 PR 13-SEP-2002; 2002US-0410566P.  
 PR 24-SEP-2002; 2002US-0413482P.  
 PR 25-SEP-2002; 2002US-0413890P.  
 PR 08-NOV-2002; 2002US-0424964P.  
 PR 13-NOV-2002; 2002US-0426222P.  
 XX  
 PA (INCY-) INCYTE CORP.  
 XX  
 PI Billiot VS, Khare R, Emerling BM, Kable AE, Tran UK, Jin P,  
 PI Becha SD, Margulis JP, Swarnakar A, Chawla NK, Ramkumar J,  
 PI Hafalia AJA, Lee SY, Jiang X, Jackson AA, Richardson TW, Blake JJ,  
 PI Wang JT, Chien D, Yang YG;  
 XX  
 DR WPI; 2004-191795/18.  
 DR P-PSDB; ADM80784.  
 XX  
 PT New cell adhesion and extracellular matrix proteins, useful in  
 PT diagnosing, treating and preventing immune, neurological, developmental,  
 PT connective tissue and cell proliferative disorders including cancer.  
 XX  
 PS Claim 5; SEQ ID NO 55; 272pp; English.  
 XX  
 CC The present sequence encodes a human cell adhesion and extracellular  
 CC matrix protein designated CADECM. CADECM sequences has neuroprotective,  
 CC cytostatic and anorectic activities. The CADECM polypeptides and  
 CC polynucleotides are useful in diagnosing, treating and preventing immune,  
 CC neurological, developmental, connective tissue and cell proliferative  
 CC disorders including cancer, e.g. breast, prostate, ovarian, lung or colon  
 CC cancer, obesity and Tangier disease.  
 XX  
 SQ Sequence 549 BP; 134 A; 203 C; 112 G; 100 T; 0 U; 0 Other;  
 XX  
 Query Match 7.2%; Score 49; DB 12; Length 549;  
 Best Local Similarity 79.5%; Pred. No. 0.0088;  
 Matches 59; Conservative 0; Mismatches 15; Indels 0; Gaps 0;  
 QY 1 ATGGGACCTGGATTGTTGCTGCTCTCGGAGACGCTTCAGTAGCTGTGAGT 60  
 DB 36 ATGGGACCTGGATTGTTGCTGCTCTCGGAGACGCTTCAGTAGCTGTGAGT 95  
 QY 61 AAAATTCCAAAT 73  
 DB 96 ACCCCTTGAAGT 108  
 Db  
 RESULT 6  
 ADM80827  
 ID ADM80827 standard; cDNA; 623 BP.  
 XX  
 AC ADM80827;  
 XX  
 DT 03-JUN-2004 (first entry)  
 XX  
 DE Human CADECM-14 encoding cDNA SEQ ID NO:56.  
 XX  
 XX human; cell adhesion and extracellular matrix protein; CADECM;  
 XX neuroprotective; cytostatic; anorectic; immune disorder;  
 XX neurological disorder; developmental disorder;  
 XX connective tissue disorder; cell proliferative disorder; cancer; obesity;  
 XX Tangier disease; gene; ss.  
 XX  
 XX Homo sapiens.  
 XX  
 XX OS  
 XX Key Location/Qualifiers  
 XX FT CDS 36..563

PT /\*tag= a  
 FT /product= "CADECM-14"  
 PN WO2004015396-A2.  
 XX  
 PD 19-FEB-2004.  
 XX  
 PF 12-AUG-2003; 2003WO-US025418.  
 XX  
 PR 13-AUG-2002; 2002US-0403781P.  
 PR 30-AUG-2002; 2002US-0407034P.  
 PR 13-SEP-2002; 2002US-0410566P.  
 PR 24-SEP-2002; 2002US-0413482P.  
 PR 25-SEP-2002; 2002US-0413890P.  
 PR 08-NOV-2002; 2002US-0424964P.  
 PR 13-NOV-2002; 2002US-0426222P.  
 XX  
 PA (INCY-) INCYTE CORP.  
 XX  
 PI Billiot VS, Khare R, Emerling BM, Kable AE, Tran UK, Jin P,  
 PI Becha SD, Margulis JP, Swarnakar A, Chawla NK, Ramkumar J,  
 PI Hafalia AJA, Lee SY, Jiang X, Jackson AA, Richardson TW, Blake JJ,  
 PI Wang JT, Chien D, Yang YG;  
 XX  
 DR WPI; 2004-191795/18.  
 DR P-PSDB; ADM80785.  
 XX  
 PT New cell adhesion and extracellular matrix proteins, useful in  
 PT diagnosing, treating and preventing immune, neurological, developmental,  
 PT connective tissue and cell proliferative disorders including cancer.  
 XX  
 PS Claim 5; SEQ ID NO 56; 272pp; English.  
 XX  
 CC The present sequence encodes a human cell adhesion and extracellular  
 CC matrix protein designated CADECM. CADECM sequences has neuroprotective,  
 CC cytostatic and anorectic activities. The CADECM polypeptides and  
 CC polynucleotides are useful in diagnosing, treating and preventing immune,  
 CC neurological, developmental, connective tissue and cell proliferative  
 CC disorders including cancer, e.g. breast, prostate, ovarian, lung or colon  
 CC cancer, obesity and Tangier disease.  
 XX  
 SQ Sequence 623 BP; 150 A; 228 C; 128 G; 117 T; 0 U; 0 Other;  
 XX  
 Query Match 7.2%; Score 49; DB 12; Length 623;  
 Best Local Similarity 79.5%; Pred. No. 0.0092;  
 Matches 58; Conservative 0; Mismatches 15; Indels 0; Gaps 0;  
 QY 1 ATGGGACCTGGATTGTTGCTGCTCTCGGAGACGCTTCAGTAGCTGTGAGT 60  
 DB 36 ATGGGACCTGGATTGTTGCTGCTCTCGGAGACGCTTCAGTAGCTGTGAGT 95  
 QY 61 AAAATTCCAAAT 73  
 DB 96 ACCCCTTGAAGT 108  
 Db  
 RESULT 7  
 ADB59026  
 ID ADB59026 standard; DNA; 556 BP.  
 XX  
 AC ADB59026;  
 XX  
 DT 04-DEC-2003 (first entry)  
 XX  
 DE Toxicity-related gene, SEQ ID 4052.  
 XX  
 XX Toxic; toxin; gene expression profile; hepatotoxicity; liver;  
 XX drug screening; toxicity assay; ds.  
 XX  
 XX Unidentified.  
 XX  
 XX OS  
 XX Key Location/Qualifiers  
 XX FT CDS WO2003064624-A2.

PD 07-AUG-2003.  
XX  
PF 31-JAN-2003; 2003MO-US003194.  
PR 31-JAN-2002; 2002US-00060087.  
PR 15-MAR-2002; 2002US-0364045P.  
PR 15-MAR-2002; 2002US-0364055P.  
PR 30-DEC-2002; 2002US-0436643P.  
XX  
PA (GENE-) GENE LOGIC INC.  
PI Mendrick D, Porter M, Johnson K, Higgs B, Caestele A, Elashoff M;  
DR WPI; 2003-689530/65.  
XX  
PT Predicting a toxic effect of a compound, useful in identifying toxicity  
PT markers in liver tissues or cells for drug screening and toxicity assays,  
PT comprises preparing gene expression profile of tissue or cells exposed to  
PT the compound.  
XX  
PS Claim 1; SEQ ID NO 4052; 1156bp; English.  
XX  
CC The present invention relates to a method for predicting a toxic effect  
CC of a compound. The method comprises preparing a gene expression profile  
CC of a tissue or cell sample exposed to the compound, and comparing the  
CC gene expression profile to a database comprising SEQ ID 1-4925, where  
CC differential expression of the gene indicates at least one toxic effect.  
CC The method is useful for predicting at least one toxic effect of a  
CC compound, predicting hepatotoxicity or the progression of a toxic effect  
CC of a compound, identifying an agent that modulates the onset or  
CC progression of a toxic response, predicting the cellular pathways that a  
CC compound modulates in a cell, and identifying an agent that modulates at  
CC least one activity of a protein. The method and compositions of the  
CC present invention using a database of genes having liver toxin-induced  
CC differential expression, are useful in identifying toxicity markers in  
CC liver tissues or cells for drug screening and toxicity assays. Note: The  
CC sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 556 BP; 121 A; 183 C; 114 G; 119 T; 0 U; 19 Other;  
XX  
Query Match 6.7%; Score 45.6; DB 10; Length 556;  
Best Local Similarity 90.6%; Pred. No. 0.069; Mismatches 5; Indels 0; Gaps 0;  
Matches 48; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
QY 1 ATGGGGAACCTGATTTTGTTCCTGCTCCTCGGAGACACCTTCACTATGCC 53  
DB 22 ATGGGGAACCTGATTTTGTTCCTGCTCCTCGGAGACACCTTGTATGCC 74  
XX  
RESULT 8  
ID ADB53782 standard; DNA; 556 BP.  
XX  
AC ADB53782;  
XX  
DT 04-DEC-2003 (first entry)  
XX  
DE Primary rat hepatocyte toxicity modelling related gene SEQ ID NO:4324.  
XX  
KM toxic effect; gene expression profile; hepatotoxicity; diagnostic marker;  
KM toxicity marker; toxicity progression; drug screening;  
KM primary rat hepatocyte toxicity modelling; gene; ds.  
XX  
OS Rattus norvegicus.  
XX  
PN MO2003065993-A2.  
XX  
PD 14-AUG-2003.  
XX  
PF 04-FEB-2003; 2003MO-US003482.  
XX

PR 04-FEB-2002; 2002US-0353171P.  
PR 13-MAR-2002; 2002US-0363534P.  
PR 08-APR-2002; 2002US-0370248P.  
PR 10-APR-2002; 2002US-0371134P.  
PR 10-APR-2002; 2002US-0371135P.  
PR 10-APR-2002; 2002US-0371150P.  
PR 11-APR-2002; 2002US-0371413P.  
PR 19-APR-2002; 2002US-0373601P.  
PR 19-APR-2002; 2002US-0373602P.  
PR 22-APR-2002; 2002US-0374139P.  
PR 08-MAY-2002; 2002US-0378370P.  
PR 09-MAY-2002; 2002US-0378652P.  
PR 09-MAY-2002; 2002US-0378653P.  
PR 09-MAY-2002; 2002US-0378665P.  
PR 09-JUL-2002; 2002US-0394230P.  
PR 09-JUL-2002; 2002US-0394253P.  
PR 04-SEP-2002; 2002US-0407688P.  
PR 28-JAN-2003; 2003US-0442900P.  
XX  
PA (GENE-) GENE LOGIC INC.  
XX  
PI Mendrick D, Porter M, Johnson K, Higgs B, Caestele A, Orr M;  
PI Elashoff M;  
XX  
DR WPI; 2003-731472/69.  
XX  
PT Determining if a compound induces a toxic effect on a tissue or cell, for  
PT identifying hepatotoxic compounds, comprises comparing a gene expression  
PT profile of a tissue or cell sample to a database of Tox mean and non-Tox  
PT mean values.  
XX  
PS Claim 44; SEQ ID NO 4324; 874bp; English.  
XX  
CC The present invention describes a method for determining whether a  
CC compound induces a toxic effect on a tissue or cell. The method comprises  
CC preparing a gene expression profile of a tissue or cell sample exposed to  
CC the compound, and comparing the gene expression profile to a database  
CC comprising data or information on the Tox mean and non-Tox mean value.  
CC The method is useful for predicting or identifying at least one toxic  
CC effect, particularly hepatotoxicity, of a test or unknown compound. The  
CC genes listed in the specification are useful as diagnostic or toxicity  
CC markers for the prediction or identification of the physiological state  
CC of tissue or cell sample that has been exposed to a compound, or to  
CC identify or predict the toxic effects of a compound or an agent. These  
CC may also be used as markers for monitoring toxicity progression or for  
CC drug screening. The present sequence represents a primary rat hepatocyte  
CC toxicity modelling related gene sequence from the present invention.  
XX  
SQ Sequence 556 BP; 121 A; 183 C; 114 G; 119 T; 0 U; 19 Other;  
XX  
Query Match 6.7%; Score 45.6; DB 10; Length 556;  
Best Local Similarity 90.6%; Pred. No. 0.069; Mismatches 5; Indels 0; Gaps 0;  
Matches 48; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
QY 1 ATGGGGAACCTGATTTTGTTCCTGCTCCTCGGAGACACCTTCACTATGCC 53  
DB 22 ATGGGGAACCTGATTTTGTTCCTGCTCCTCGGAGACACCTTGTATGCC 74  
XX  
RESULT 9  
ID AA250832 standard; DNA; 476 BP.  
XX  
AC AA250832;  
XX  
DT 31-MAY-2000 (first entry)  
XX  
DE Rat amelogenin gene (A4).  
XX  
KM Amelogenin; splice variant; rat; (A4); chondrogenesis; osteogenesis;  
KM chondrogenic inducing molecule; CIM; cartilage growth; osteopathic;  
KM extracellular matrix protein; tooth enamel; enamel mineralisation;  
KM ameloblast; bone regeneration; composite cell construct; ds.  
XX

```

XX OS Rattus sp.
XX
XX Key Location/Qualifiers
XX FT exon 1..36
XX FT exon /*tag= a
XX FT exon /partial
XX FT exon /number= 1
XX FT exon 37..101
XX FT exon /*tag= b
XX FT exon /number= 2
XX CDS 48..317
XX FT exon /*tag= c
XX FT exon /product= "Rat amelogenin protein"
XX FT exon /transl_except= (pos:69..71, aa:Gly)
XX FT exon 48..95
XX FT exon /*tag= d
XX FT exon 96..314
XX FT exon /*tag= e
XX FT exon /label= Mature_rat_amelogenin_protein
XX FT exon 102..149
XX FT exon /*tag= f
XX FT exon /number= 3
XX FT exon 150..191
XX FT exon /*tag= g
XX FT exon /number= 4
XX FT exon 192..236
XX FT exon /*tag= h
XX FT exon /number= 5
XX FT exon 237..311
XX FT exon /*tag= i
XX FT exon /number= 6d
XX FT exon /note= "Comprises of gene segments 6a, b, c and d"
XX FT exon 312..317
XX FT exon /*tag= j
XX FT exon /number= 7
XX FT exon /note= "Includes the stop codon"
XX PN W0200006734-A1.
XX
XX PD 10-FEB-2000.
XX
XX PE 29-JUL-1999; 99WO-US017242.
XX
XX PR 29-JUL-1998; 98US-0094489P.
XX
XX PA (NOUN ) UNIV NORTHWESTERN.
XX
XX PI Vels A, Nebgen DR;
XX
XX DR WPI; 2000-205464/18.
XX
XX DR P-PSDB; AAY45074.
XX
XX PT Novel amelogenin polypeptides and polynucleotides, useful for enhancing
XX PT bone generation in mammals and synthesizing bone matrix or articular
XX PT surfaces at implant sites.
XX
XX PS Example 2; Fig 11B; 79pp; English.
XX
XX CC The present DNA sequence is the full-length rat amelogenin gene (A4),
XX CC comprising exons 1-7, including the exon segment 6d. It is derived from
XX CC the rat incisor odontoblast-pulp cDNA library. The splice variants of
XX CC this gene functions as an osteogenic or chondrogenic inducing molecule
XX CC (CIM), which is useful for enhancing bone or cartilage growth. It has
XX CC osteoplastic activity. Amelogenin belongs to the family of extracellular
XX CC matrix proteins, in developing tooth enamel, that are produced by the
XX CC ameloblasts and plays a role in enamel mineralisation. Chondrogenic or
XX CC osteogenic inducing amelogenin molecules are useful to induce
XX CC differentiation of cells to the osteogenic and chondrogenic phenotypes
XX CC and can be used in a composite cell construct for bone and cartilage
XX CC regeneration. The polynucleotides can be employed to produce the
XX CC polypeptides by recombinant techniques

```

```

SQ Sequence 476 BP; 155 A; 106 C; 98 G; 117 T; 0 U; 0 Other;
Query Match 6.6%; Score 45; DB 3; Length 476;
Best Local Similarity 90.6%; Pred. No. 0.095;
Matches 48; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1 ATGGGACCTGGATTGTTTTCCTGCTCTGAGAGACCTGATGATGCC 53
DB 48 ATGGGACCTGGATTGTTTTCCTGCTCTGAGAGACCTGATGATGCC 100
RESULT 10
ADM80836
ID ADM80836 standard; cDNA; 318 BP.
AC ADM80836;
DE 03-JUN-2004 (first entry)
XX Human CADECM-23 encoding cDNA SEQ ID NO:65.
XX
XX KW human; cell adhesion and extracellular matrix protein; CADECM;
XX KW neuroprotective; cytoskeletal; anorectic; immune disorder;
XX KW neurological disorder; developmental disorder;
XX KW connective tissue disorder; cell proliferative disorder; cancer; obesity;
XX KW Tangier disease; gene; ss.
XX
XX OS Homo sapiens.
XX
XX FH Key Location/Qualifiers
XX FT CDS 36..260
XX FT exon /*tag= a
XX FT exon /product= "CADECM-23"
XX
XX PN W02004015396-A2.
XX
XX PD 19-FEB-2004.
XX
XX PE 12-AUG-2003; 2003WO-US025418.
XX
XX PR 13-AUG-2002; 2002US-0403781P.
XX PR 30-AUG-2002; 2002US-0407034P.
XX PR 13-SEP-2002; 2002US-0410566P.
XX PR 24-SEP-2002; 2002US-0413482P.
XX PR 25-SEP-2002; 2002US-0413890P.
XX PR 08-NOV-2002; 2002US-0424804P.
XX PR 13-NOV-2002; 2002US-0426222P.
XX
XX PA (INCY-) INCYTE CORP.
XX
XX PI Elliott VS, Khare R, Emerling BM, Kable AE, Tran UK, Jin P,
XX PI Becha SD, Margulis JP, Swarnakar A, Chewla NK, Ramkumar J,
XX PI Hafalala AJA, Lee SY, Ujang X, Jackson AA, Richardson TW, Blake JJ,
XX PI Wang JT, Chien D, Yang YG;
XX
XX DR WPI; 2004-191795/18.
XX
XX DR P-PSDB; ADM80794.
XX
XX PT New cell adhesion and extracellular matrix proteins, useful in
XX PT diagnosing, treating and preventing immune, neurological, developmental,
XX PT connective tissue and cell proliferative disorders including cancer.
XX
XX PS Claim 5; SEQ ID NO 65; 272pp; English.
XX
XX CC The present sequence encodes a human cell adhesion and extracellular
XX CC matrix protein designated CADECM. CADECM sequences has neuroprotective,
XX CC cytoskeletal and anorectic activities. The CADECM polypeptides and
XX CC polynucleotides are useful in diagnosing, treating and preventing immune,
XX CC neurological, developmental, connective tissue and cell proliferative
XX CC disorders including cancer, e.g. breast, prostate, ovarian, lung or colon
XX CC cancer, obesity and Tangier disease.
XX
XX SQ Sequence 318 BP; 87 A; 81 C; 75 G; 75 T; 0 U; 0 Other;

```

Query Match 6.6%; Score 44.8; DB 12; Length 318;  
Best Local Similarity 87.5%; Pred. No. 0.095; Mismatches 7; Indels 0; Gaps 0;  
Matches 49; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 ATGGGACCTGATTTTGTGCTGCTCTGAGAGACCTTTCAGTATGCTGT 56  
DB 36 ATGGGACCTGATTTTATTTGCTGCTCTGAGAGACCTTTCAGTATGCTGT 91

RESULT 11  
AAZ07020 standard; DNA; 750 BP.

AC AAZ07020;  
XX  
XX 15-NOV-1999 (first entry)  
XX  
XX Amelogenin X nucleotide sequence.  
XX  
XX  
XX Amelogenin X; Amelogenin Y; genotype; 5' nuclease amplification;  
XX multiple allelic site; apolipoprotein E; apob; coronary artery disease;  
XX Alzheimer's disease; ds.  
XX  
XX Unidentified.  
XX  
XX MO9940226-A2.  
XX  
XX 12-AUG-1999.  
XX  
XX 08-JAN-1999; 99MO-US000499.  
XX  
XX 04-FEB-1998; 98US-00018595.  
XX  
XX (PEKE ) PERKIN-ELMER CORP.  
XX  
XX Livak KJ, Goodsaiaid F;  
XX  
XX MPI; 1999-539985/45.  
XX  
XX 5' nuclease amplification assay using fluorescence-quencher probes for  
XX determination of a genotype at multiple allelic sites.  
XX  
XX Disclosure; Fig 10; 95pp; English.  
XX  
XX The present invention describes first and second sets of fluorescer-  
XX quencher probes used simultaneously in a 5' nuclease assay to identify  
XX which members of a first or second set of substantially homologous  
XX sequences are present in a DNA sample. The method can be used to genotype  
XX a sample of genomic DNA at two or more different allelic sites.  
XX Generating a fluorescence spectrum and signature for each genotype, which  
XX uniquely reflects the assay's inherent inefficiency for that genotype  
XX given the particular conditions, probes and primers used, the genotype of  
XX unknown sequences can be determined. The assay was shown to be useful for  
XX determining apob genotypes. The assay can be used as a diagnostic tool  
XX for assessing the risk for coronary artery disease and/or late-onset  
XX Alzheimer's disease. Using the 5' nuclease assay of the invention it is  
XX possible to determine a genotype at two or more allelic sites in a single  
XX reaction. This approach is much faster than previous approaches to  
XX genotyping genes having two or more allelic sites, such as the  
XX apolipoprotein E gene. A key advantage of the method for determining the  
XX genotype of a sample of DNA at multiple allelic sites is that it does not  
XX rely on 5' nuclease assay working with 100% efficiency to distinguish  
XX between substantially homologous sequences such as alleles. The present  
XX sequence represent the nucleotide sequence for amelogenin X, which is  
XX used in the exemplification of the present invention

SO Sequence 750 BP; 191 A; 260 C; 146 G; 153 T; 0 U; 0 Other;

Query Match 6.6%; Score 44.8; DB 2; Length 750;  
Best Local Similarity 87.5%; Pred. No. 0.12; Mismatches 7; Indels 0; Gaps 0;  
Matches 49; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 ATGGGACCTGATTTTGTGCTGCTCTGAGAGACCTTTCAGTATGCTGT 56  
DB 69 ATGGGACCTGATTTTATTTGCTGCTCTGAGAGACCTTTCAGTATGCTGT 124

RESULT 12  
AAZ07018 standard; DNA; 793 BP.

AC AAZ07018;  
XX  
XX 15-NOV-1999 (first entry)  
XX  
XX Amelogenin X nucleotide sequence.  
XX  
XX  
XX Amelogenin X; Amelogenin Y; genotype; 5' nuclease amplification;  
XX multiple allelic site; apolipoprotein E; apob; coronary artery disease;  
XX Alzheimer's disease; ss.  
XX  
XX Unidentified.  
XX  
XX MO9940226-A2.  
XX  
XX 12-AUG-1999.  
XX  
XX 08-JAN-1999; 99MO-US000499.  
XX  
XX 04-FEB-1998; 98US-00018595.  
XX  
XX (PEKE ) PERKIN-ELMER CORP.  
XX  
XX Livak KJ, Goodsaiaid F;  
XX  
XX MPI; 1999-539985/45.  
XX  
XX 5' nuclease amplification assay using fluorescence-quencher probes for  
XX determination of a genotype at multiple allelic sites.  
XX  
XX Disclosure; Fig 8A; 95pp; English.  
XX  
XX The present invention describes first and second sets of fluorescer-  
XX quencher probes used simultaneously in a 5' nuclease assay to identify  
XX which members of a first or second set of substantially homologous  
XX sequences are present in a DNA sample. The method can be used to genotype  
XX a sample of genomic DNA at two or more different allelic sites.  
XX Generating a fluorescence spectrum and signature for each genotype, which  
XX uniquely reflects the assay's inherent inefficiency for that genotype  
XX given the particular conditions, probes and primers used, the genotype of  
XX unknown sequences can be determined. The assay was shown to be useful for  
XX determining apob genotypes. The assay can be used as a diagnostic tool  
XX for assessing the risk for coronary artery disease and/or late-onset  
XX Alzheimer's disease. Using the 5' nuclease assay of the invention it is  
XX possible to determine a genotype at two or more allelic sites in a single  
XX reaction. This approach is much faster than previous approaches to  
XX genotyping genes having two or more allelic sites, such as the  
XX apolipoprotein E gene. A key advantage of the method for determining the  
XX genotype of a sample of DNA at multiple allelic sites is that it does not  
XX rely on 5' nuclease assay working with 100% efficiency to distinguish  
XX between substantially homologous sequences such as alleles. The present  
XX sequence represent the nucleotide sequence for amelogenin X, which is  
XX used in the exemplification of the present invention

SO Sequence 793 BP; 215 A; 266 C; 150 G; 162 T; 0 U; 0 Other;

Query Match 6.6%; Score 44.8; DB 2; Length 793;  
Best Local Similarity 87.5%; Pred. No. 0.12; Mismatches 7; Indels 0; Gaps 0;  
Matches 49; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 ATGGGACCTGATTTTGTGCTGCTCTGAGAGACCTTTCAGTATGCTGT 56  
DB 69 ATGGGACCTGATTTTATTTGCTGCTCTGAGAGACCTTTCAGTATGCTGT 124

RESULT 13  
AAZ07019  
ID AAZ07019 standard; DNA; 802 BP.  
XX  
AC AAZ07019;  
XX  
DT 15-NOV-1999 (first entry)  
XX  
DE Amelogenin Y nucleotide sequence.  
XX  
KM Amelogenin X; Amelogenin Y; genotype; 5' nuclease amplification;  
XX multiple allele site; apolipoprotein E; apoB; coronary artery disease;  
XX Alzheimer's disease; ss.  
XX  
OS Unidentified.  
XX  
PN WO940226-A2.  
XX  
PD 12-AUG-1999.  
XX  
PF 08-JAN-1999; 99MO-US000499.  
XX  
PR 04-FEB-1998; 98US-00018595.  
XX  
PA (PEKE ) PERKIN-ELMER CORP.  
XX  
PI Liyak KJ, Goodaaid F;  
XX  
DR WPI; 1999-539985/45.  
XX  
PT 5' nuclease amplification assay using fluorescence-quencher probes for  
XX determination of a genotype at multiple allelic sites.  
XX  
PS Disclosure; Fig 8B; 95pp; English.  
XX  
XX The present invention describes first and second sets of fluorescer-  
CC quencher probes used simultaneously in a 5' nuclease assay to identify  
CC which members of a first or second set of substantially homologous  
CC sequences are present in a DNA sample. The method can be used to genotype  
CC a sample of genomic DNA at two or more different allelic sites.  
CC Generating a fluorescence spectrum and signature for each genotype, which  
CC uniquely reflects the assay's inherent inefficiency for that genotype  
CC given the particular conditions, probes and primers used, the genotype of  
CC unknown sequences can be determined. The assay was shown to be useful for  
CC determining apoe genotypes. The assay can be used as a diagnostic tool  
CC for assessing the risk for coronary artery disease and/or late-onset  
CC Alzheimer's disease. Using the 5' nuclease assay of the invention it is  
CC possible to determine a genotype at two or more allelic sites in a single  
CC reaction. This approach is much faster than previous approaches to  
CC genotyping genes having two or more allelic sites, such as the  
CC apolipoprotein E gene. A key advantage of the method for determining the  
CC genotype of a sample of DNA at multiple allelic sites is that it does not  
CC rely on 5' nuclease assay working with 100% efficiency to distinguish  
CC between substantially homologous sequences such as alleles. The present  
CC sequence represent the nucleotide sequence for amelogenin Y, which is  
CC used in the exemplification of the present invention  
XX  
SQ Sequence 802 BP; 216 A; 258 C; 155 G; 173 T; 0 U; 0 Other;  
XX  
Query Match 6.4%; Score 43.2; DB 2; Length 802;  
Best Local Similarity 85.7%; Pred. No. 0.33;  
Matches 48; Conservative 0; Mismatches 8; Indels 0; Gaps 0;  
XX  
QY 1 ATGGGACCTGGATTGTTGCTGCTGCTGGAGACCTTCAGTATGCTCT 56  
DB 69 ATGGGACCTGGATTGTTGCTGCTGCTGGAGACCTTCAGTATGCTCT 124  
XX  
RESULT 14  
ADQ22977  
XX  
ID ADQ22977 standard; DNA; 852 BP.  
XX  
AC ADQ22977;  
XX

XX  
DT 26-AUG-2004 (first entry)  
XX  
DE Human soft tissue sarcoma-upregulated DNA - SEQ ID 5797.  
XX  
KM Human soft tissue sarcoma; cytosolic; gene therapy; vaccine; screening; human;  
XX da.  
XX  
OS Homo sapiens.  
XX  
PN WO2004048938-A2.  
XX  
PD 10-JUN-2004.  
XX  
PF 26-NOV-2003; 2003WO-US038193.  
XX  
PR 26-NOV-2002; 2002US-0429739P.  
XX  
PA (PROT-) PROTEIN DESIGN LABS INC.  
XX  
PI Aziz N, Ginsburg WM, Zlotnick A;  
XX  
DR WPI; 2004-441208/41.  
XX  
PT Early detection of soft tissue sarcoma comprises determining expression  
XX of a gene in a first soft tissue sample and a normal soft tissue sample  
XX and comparing the gene expression, also useful in treating soft tissue  
XX sarcoma.  
XX  
PS Example 2; SEQ ID NO 5797; 210pp; English.  
XX  
XX The invention relates to a novel method for detecting soft tissue sarcoma  
CC which comprises obtaining a first soft tissue sample from an individual  
CC and a normal soft tissue sample from the same or different individual,  
CC determining the expression of a gene in both samples and comparing the  
CC expression of the gene in both soft tissue samples, where a higher level  
CC of protein expression in the first soft tissue sample indicates the  
CC presence of soft tissue sarcoma. The method of the invention has  
CC cytosolic applications and may be useful for detecting soft tissue  
CC sarcoma, possibly via gene therapy or vaccine production. The nucleic  
CC acid sequences may be useful in diagnostic and screening applications.  
CC The current sequence is that of a human soft tissue sarcoma-upregulated  
CC DNA of the invention. The current sequence is not shown within the  
CC specification per se but was submitted in CD format by the inventor.  
XX  
SQ Sequence 852 BP; 226 A; 267 C; 169 G; 190 T; 0 U; 0 Other;  
XX  
Query Match 6.4%; Score 43.2; DB 12; Length 852;  
Best Local Similarity 85.7%; Pred. No. 0.33;  
Matches 48; Conservative 0; Mismatches 8; Indels 0; Gaps 0;  
XX  
QY 1 ATGGGACCTGGATTGTTGCTGCTGCTGGAGACCTTCAGTATGCTCT 56  
DB 69 ATGGGACCTGGATTGTTGCTGCTGCTGGAGACCTTCAGTATGCTCT 124  
XX  
RESULT 15  
ABQ88146  
XX  
ID ABQ88146 standard; cDNA; 178896 BP.  
XX  
AC ABQ88146;  
XX  
DT 18-SEP-2002 (first entry)  
XX  
DE Human osteoblast differentiation related cDNA SEQ ID NO 53.  
XX  
KM Human; osteoblast; stem cell differentiation; bone tissue deposition;  
XX osteoporosis; osteopathc; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200250301-A2.  
XX

PD 27-JUN-2002.  
XX  
XX 18-DEC-2001; 2001WO-US048276.  
PF  
XX 18-DEC-2000; 2000US-0255882P.  
PR  
XX 24-APR-2001; 2001US-0285691P.  
PR  
XX (GENE-) GENE LOGIC INC.  
PA (PROC ) PROCTER & GAMBLE CO.  
XX  
XX JI D, Axelrod DW, Cook JS, Jaiswal N, Einstein R, Houghton A;  
PI Mertz LJ  
XX  
XX MPI; 2002-557663/59.  
DR  
XX  
XX Use of genes and their expression profiles associated with osteoblast  
PT differentiation for screening modulators bone formation, for diagnosing  
PT or treating e.g. osteoporosis, or as markers for the differentiation  
PT process.  
XX  
PS Claim 1, SEQ ID NO 53; 78bp + Sequence Listing; English.  
XX  
XX The invention relates to genes and their expression profiles are used  
CC for: (a) screening modulators of precursor stem cell differentiation into  
CC osteoblasts, or bone tissue deposition; (b) diagnosing abnormal  
CC deposition of bone tissue, abnormal rate of osteoblast formation or  
CC osteoporosis; or (c) treating or monitoring treatment of the conditions  
CC cited in (b), or monitoring the progression of bone tissue deposition.  
CC Specific conditions include postmenopausal osteoporosis, glucocorticoid  
CC osteoporosis or male osteoporosis, osteopenia, osteodystrophy, drug-  
CC induced abnormalities in bone formation or bone loss, conditions that  
CC involve altered bone metabolism (e.g. idiopathic juvenile osteoporosis),  
CC skeletal disease linked to breast cancer, mastocytosis, Fanconi syndrome  
CC or fibrous dysplasia. The present sequence is that of an osteoblast  
CC differentiation associated CDNA marker of the invention. Note: The  
CC sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 178896 BP; 50288 A; 36724 C; 39084 G; 52800 T; 0 U; 0 Other;  
Query Match 6.3%; Score 43; DB 6; Length 178896;  
Best Local Similarity 46.5%; Pred. No. 1.8;  
Matches 139; Conservative 0; Mismatches 160; Indels 0; Gaps 0;  
QY 165. CACAAATGTCAGTGTGTTAAGAAACATTTTAAATCTTGATCAGCAAGATTTTAAACA 224  
DB 73412 CAAAAATGACAAATGGAATCATCATCAAGTTAAAGCTTCTGCACAGCCAGATACAA 73471  
QY 225 AAAACATTGTCAGTGTGCTTCAACCATATTCATTCAGTATAGCCAGAGTGGGGGTG 284  
DB 73472 CCAACAAAGAGAGAGCAACCCAGAAATAGAGAAATATCTGTAACTACCATCTG 73531  
QY 285 GCTGAAACCTGTAATTTTAAAGCAAGAACTCAGAAATCTTGAAACATTAATT 344  
DB 73532 ACAAGGATTAATTAACCAAGATATTAAGAGCTCAACAACCTCTATAGAAAAATCTA 73591  
QY 345 ATATATCTCATCTCTAGATGAAAAATTTTGAAGGACCTTGAAGGCTCCAGAAA 404  
DB 73592 ATATATCTGAATTTAAAAAATGGCAAAAGATTGAATAGACATTTCTCAAAAAGACAT 73651  
QY 405 AAGTGCTCAATCAGTCTTTTAAGTACTAACAATAATGCCAGTTTCTTAATAACCAATT 463  
DB 73652 CAGAGGCAAAAGAGGATGATATAGATTAATGAATAATCACAACCTATATATCAATAT 73710

Search completed: February 8, 2005, 13:57:14  
Job time : 905.049 secs

This Page Blank (uspto)





; Sequence 3, Application US/09744128  
; Patent No. 6677306  
; GENERAL INFORMATION:  
; APPLICANT: Veis et al.  
; TITLE OF INVENTION: Chondrogenic and Osteogenic Inducing Molecule  
; FILE REFERENCE: 27636/36983  
; CURRENT APPLICATION NUMBER: US/09/744,128  
; CURRENT FILING DATE: 2001-05-16  
; PRIOR APPLICATION NUMBER: PCT/US99/17342  
; PRIOR FILING DATE: 1999-07-29  
; PRIOR APPLICATION NUMBER: 60/094,489  
; PRIOR FILING DATE: 1998-07-29  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: PatentIn 3.1  
; SEQ ID NO 3  
; LENGTH: 476  
; TYPE: DNA  
; ORGANISM: Artificial sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: PCR Product  
US-09-744-128-3

Query Match 6.6%; Score 45; DB 4; Length 476;  
Best Local Similarity 90.6%; Pred. No. 0.0037;  
Matches 48; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ATGGGACCTGGATTGTTGCTGCTGCTGGAGACAGCTTCAGTATGCC 53  
DB 48 ATGGGACCTGGATTGTTGCTGCTGCTGGAGACAGCTTCAGTATGCC 100

RESULT 3  
US-09-018-595B-3  
; Sequence 3, Application US/09018595B  
; Patent No. 5962233  
; GENERAL INFORMATION:  
; APPLICANT: Perkin-Elmer Corporation,  
; APPLICANT: Applied Biosystems Division  
; TITLE OF INVENTION: DETERMINATION OF GENOTYPE OF  
; TITLE OF INVENTION: AMPLIFICATION PRODUCTS AT MULTIPLE ALLELIC SITES  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: David J. Weitz,  
; ADDRESSEE: Wilson Sonsini Goodrich & Rosati  
; STREET: 650 Page Mill Road  
; CITY: Palo Alto  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94304-1050  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 inch diskette  
; COMPUTER: IBM compatible  
; OPERATING SYSTEM: Microsoft Windows 95  
; SOFTWARE: Wordperfect for windows 6.0,  
; SOFTWARE: ASCII (DOS) TEXT format  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/018,595B  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: David J. Weitz  
; REGISTRATION NUMBER: 38,362  
; REFERENCE/DOCKET NUMBER: PELM-744  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (650) 493-9300  
; TELEFAX: (650) 493-6811  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 500 nucleotides  
; TYPE: nucleic acid

; STRANDEDNESS: double  
; TOPOLOGY: linear  
US-09-018-595B-3

Query Match 6.6%; Score 44.8; DB 2; Length 500;  
Best Local Similarity 87.5%; Pred. No. 0.0043;  
Matches 49; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 ATGGGACCTGGATTGTTGCTGCTGCTGGAGACAGCTTCAGTATGCC 56  
DB 69 ATGGGACCTGGATTGTTGCTGCTGCTGGAGACAGCTTCAGTATGCC 124

RESULT 4  
US-09-324-709A-3  
; Sequence 3, Application US/09324709A  
; Patent No. 6154707  
; GENERAL INFORMATION:  
; APPLICANT: Perkin-Elmer Corporation,  
; APPLICANT: Applied Biosystems Division  
; TITLE OF INVENTION: FLUORESCENCE GENOTYPING AT MULTIPLE ALLELIC  
; TITLE OF INVENTION: SITES  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: David J. Weitz,  
; ADDRESSEE: Wilson Sonsini Goodrich & Rosati  
; STREET: 650 Page Mill Road  
; CITY: Palo Alto  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94304-1050  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 inch diskette  
; COMPUTER: IBM compatible  
; OPERATING SYSTEM: Microsoft Windows 95  
; SOFTWARE: Wordperfect for windows 6.0,  
; SOFTWARE: ASCII (DOS) TEXT format  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/324,709A  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: David J. Weitz  
; REGISTRATION NUMBER: 38,362  
; REFERENCE/DOCKET NUMBER: 16842-758  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (650) 493-9300  
; TELEFAX: (650) 493-6811  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 500 nucleotides  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
US-09-324-709A-3

Query Match 6.6%; Score 44.8; DB 3; Length 500;  
Best Local Similarity 87.5%; Pred. No. 0.0043;  
Matches 49; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 ATGGGACCTGGATTGTTGCTGCTGCTGGAGACAGCTTCAGTATGCC 56  
DB 69 ATGGGACCTGGATTGTTGCTGCTGCTGGAGACAGCTTCAGTATGCC 124

RESULT 5  
US-09-018-595B-1  
; Sequence 1, Application US/09018595B  
; Patent No. 5962233  
; GENERAL INFORMATION:

```

; APPLICANT: Perkin-Elmer Corporation,
; TITLE OF INVENTION: DETERMINATION OF GENOTYPE OF
; TITLE OF INVENTION: AMPLIFICATION PRODUCTS AT MULTIPLE ALLELIC SITES
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: David J. Weltz,
; STREET: 650 Page Mill Road
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1050
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch diskette
; OPERATING SYSTEM: Microsoft Windows 95
; SOFTWARE: Wordperfect for windows 6.0,
; SOFTWARE: ASCII (DOS) TEXT format
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/018,595B
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: David J. Weltz
; REGISTRATION NUMBER: 38,362
; REFERENCE/DOCKET NUMBER: PELM-744
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 493-9300
; TELEFAX: (650) 493-6811
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 793 nucleotides
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-09-018-595B-1
;
Query Match 6.6%; Score 44.8; DB 2; Length 793;
Best Local Similarity 87.5%; Pred. No. 0.0052;
Matches 49; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
;
Qy 1 ATGGGACCTGATTTGTTGCTGCTCGGAGCAGCCTTCAGTAGCCTGT 56
Db 69 ATGGGACCTGATTTATTTATTTGCTGCTCGGAGCAGCCTTTGGCAGCCTCT 124
;
RESULT 6
US-09-324-709A-1
; Sequence 1, Application US/09324709A
; Patent No. 6154707
; GENERAL INFORMATION:
; APPLICANT: Perkin-Elmer Corporation,
; TITLE OF INVENTION: Applied Biosystems Division
; TITLE OF INVENTION: FLUORESCENCE GENOTYPING AT MULTIPLE ALLELIC
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: David J. Weltz,
; STREET: 650 Page Mill Road
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1050
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch diskette
; OPERATING SYSTEM: Microsoft Windows 95
; SOFTWARE: Wordperfect for windows 6.0,
; SOFTWARE: Wordperfect for windows 6.0,
```

```

; SOFTWARE: ASCII (DOS) TEXT format
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/324,709A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: David J. Weltz
; REGISTRATION NUMBER: 38,362
; REFERENCE/DOCKET NUMBER: 16842-758
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 493-9300
; TELEFAX: (650) 493-6811
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 793 nucleotides
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-09-324-709A-1
;
Query Match 6.6%; Score 44.8; DB 3; Length 793;
Best Local Similarity 87.5%; Pred. No. 0.0052;
Matches 49; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
;
Qy 1 ATGGGACCTGATTTGTTGCTGCTCGGAGCAGCCTTCAGTAGCCTGT 56
Db 69 ATGGGACCTGATTTATTTATTTGCTGCTCGGAGCAGCCTTTGGCAGCCTCT 124
;
RESULT 7
US-09-018-595B-2
; Sequence 2, Application US/09018595B
; Patent No. 596223
; GENERAL INFORMATION:
; APPLICANT: Perkin-Elmer Corporation,
; TITLE OF INVENTION: Applied Biosystems Division
; TITLE OF INVENTION: DETERMINATION OF GENOTYPE OF
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: David J. Weltz,
; STREET: 650 Page Mill Road
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1050
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch diskette
; OPERATING SYSTEM: Microsoft Windows 95
; SOFTWARE: Wordperfect for windows 6.0,
; SOFTWARE: ASCII (DOS) TEXT format
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/018,595B
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: David J. Weltz
; REGISTRATION NUMBER: 38,362
; REFERENCE/DOCKET NUMBER: PELM-744
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 493-9300
; TELEFAX: (650) 493-6811
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 802 nucleotides
```

TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-018-5958-2

Query Match 6.4%; Score 43.2; DB 2; Length 802;  
Best Local Similarity 85.7%; Pred. No. 0.015;  
Matches 48; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 ATGGGACCTGATTTGTTGCTGCTGCTGAGACCTTCAGTATGCTGT 56  
DB 69 ATGGGACCTGATTTGTTGCTGCTGCTGAGACCTTCAGTATGCTGT 124

RESULT 8  
US-09-324-709A-2  
Sequence 2, Application US/09324709A  
Patent No. 6154707

GENERAL INFORMATION:  
APPLICANT: Perkin-Elmer Corporation,  
APPLICANT: Applied Biosystems Division  
TITLE OF INVENTION: FLUORESCENCE GENOTYPING AT MULTIPLE ALLELIC  
TITLE OF INVENTION: SITES  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:

ADDRESSEE: David J. Weltz,  
ADDRESSEE: Wilson Sonsini Goodrich & Rosati  
STREET: 650 Page Mill Road  
CITY: Palo Alto  
STATE: California  
COUNTRY: USA  
ZIP: 94304-1050

COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch diskette  
COMPUTER: IBM compatible  
OPERATING SYSTEM: Microsoft Windows 95  
SOFTWARE: Wordperfect for windows 6.0,  
SOFTWARE: ASCII (DOS) TEXT format  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/324,709A

FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: David J. Weltz

REGISTRATION NUMBER: 38,362

REFERENCE/DOCKET NUMBER: 16842-758

TELECOMMUNICATION INFORMATION:

TELEPHONE: (650) 493-9300

TELEFAX: (650) 493-6811

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 802 nucleotides

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

US-09-324-709A-2

Query Match 6.4%; Score 43.2; DB 3; Length 802;  
Best Local Similarity 85.7%; Pred. No. 0.015;  
Matches 48; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 ATGGGACCTGATTTGTTGCTGCTGCTGAGACCTTCAGTATGCTGT 56  
DB 69 ATGGGACCTGATTTGTTGCTGCTGCTGAGACCTTCAGTATGCTGT 124

RESULT 9  
US-09-806-708B-22/C  
Sequence 22, Application US/09806708B  
Patent No. 6784342

GENERAL INFORMATION:

APPLICANT: The University of British Columbia

TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants

FILE REFERENCE: 4810-58741

CURRENT APPLICATION NUMBER: US/09/806,708B

CURRENT FILING DATE: 2001-04-03

PRIOR APPLICATION NUMBER: US 60/147,133

PRIOR FILING DATE: 1999-08-04

NUMBER OF SEQ ID NOS: 23

SOFTWARE: PatentIn version 3.0

SEQ ID NO 22

LENGTH: 1141

TYPE: DNA

ORGANISM: Artificial sequence

FEATURE:

NAME/KEY: promoter

LOCATION: (1)..(1141)

OTHER INFORMATION: consensus sequence of A.t., L.a., and B.n. FAEI promoters

US-09-806-708B-22

Query Match 6.0%; Score 40.8; DB 4; Length 1141;  
Best Local Similarity 8.0%; Pred. No. 0.079;  
Matches 39; Conservative 213; Mismatches 233; Indels 0; Gaps 0;

QY 182 TTAAGAAACATTTTAAATCTTGATCAACAATTTTAAACAAACATTCAGTTC 241  
DB 626 WTAKGDMTRKVKVKKRDTTCYVDVADSWWYAMRCDVYTRNNYCKSYAHSY 567

QY 242 CTTCACCAATATTCATTCAGTATAGCAGAGGCGGCTGTCAGAACTGTAAT 301  
DB 566 WYMSNNAMWYRYSAWMSWABWTRNNWMSGVBWBMAGTMMBRNNNNNTDRRYW 507

QY 302 TTAACAGCAAGAACTCAGAAATTCCTGAAACATTTATATATCATCATCTCTA 361  
DB 506 WKRWARBYTTTVDSCNCAKSMWRGNMRAMKMMWAAANNDAAGMDWTYMGNTMMBR 447

QY 362 GATGAAATTTTGGAGAGGACCTTTGAAAGGCGCCAGAAAGAGTGTCAATCATGCT 421  
DB 446 AWMKMMWACRAVYCCNNNNNPAVHKKKMRWTKYKMAKCNNNBRKAMRYAMM 387

QY 422 TTAAGTACTACAAAATGCAAGTTCCTTAAACCAATTCACAGATCTAATC 481  
DB 386 YSRDTNTDMMWTSDBWHTVTDYTWBAMNNNNNNWBRCKTSMWMMMDHMTCT 327

QY 482 TCTTCTGCTCAAAAACCCCGGCTATCATTTCTCAAGTAAAGTTCAGTGTAGA 541  
DB 326 YGNNTWGSAVBMWAAASMBVTVWVCWRTYMGKTMNNNNNNRATYRTYVAM 267

QY 542 ATTTCCTCCTGAACATTTATTTGTCAGTCACTGCTAAGAGATGATCTTATGCT 601  
DB 266 CANNRYIDPAWVBKRNRYCTAYBMYVMYKHHMBWRRAHRSMMWVYKCNKYM 207

QY 602 TGTCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 661  
DB 206 VSMHYAMRYBKWABAVGCNNNMKDMAHHCATNNNNMMWYAVMHHMKKGRKAAW 147

QY 662 GTTAC 666  
DB 146 NKTAB 142

RESULT 10  
US-09-248-796A-11112  
Sequence 1112, Application US/09248796A  
Patent No. 6747137

GENERAL INFORMATION:

APPLICANT: Keith Wainrock et al

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAT

FILE REFERENCE: 107196.132

CURRENT APPLICATION NUMBER: US/09/248,796A

CURRENT FILING DATE: 1999-02-12

PRIOR APPLICATION NUMBER: US 60/074,725

US-09-248-796A-11112

Query Match 6.0%; Score 40.8; DB 4; Length 1141;  
Best Local Similarity 8.0%; Pred. No. 0.079;  
Matches 39; Conservative 213; Mismatches 233; Indels 0; Gaps 0;

QY 182 TTAAGAAACATTTTAAATCTTGATCAACAATTTTAAACAAACATTCAGTTC 241  
DB 626 WTAKGDMTRKVKVKKRDTTCYVDVADSWWYAMRCDVYTRNNYCKSYAHSY 567

QY 242 CTTCACCAATATTCATTCAGTATAGCAGAGGCGGCTGTCAGAACTGTAAT 301  
DB 566 WYMSNNAMWYRYSAWMSWABWTRNNWMSGVBWBMAGTMMBRNNNNNTDRRYW 507

QY 302 TTAACAGCAAGAACTCAGAAATTCCTGAAACATTTATATATCATCATCTCTA 361  
DB 506 WKRWARBYTTTVDSCNCAKSMWRGNMRAMKMMWAAANNDAAGMDWTYMGNTMMBR 447

QY 362 GATGAAATTTTGGAGAGGACCTTTGAAAGGCGCCAGAAAGAGTGTCAATCATGCT 421  
DB 446 AWMKMMWACRAVYCCNNNNNPAVHKKKMRWTKYKMAKCNNNBRKAMRYAMM 387

;; PRIOR FILING DATE: 1998-02-13  
;; PRIOR APPLICATION NUMBER: US 60/096,409  
;; PRIOR FILING DATE: 1998-08-13  
;; NUMBER OF SEQ ID NOS: 28208  
;; SEQ ID NO 11112  
;; LENGTH: 1686  
;; TYPE: DNA  
;; ORGANISM: Candida albicans  
US-09-248-796A-11112

Query Match 5.8%; Score 39.6; DB 4; Length 1686;  
Best Local Similarity 44.1%; Pred. No. 0.2;  
Matches 165; Conservative 0; Mismatches 209; Indels 0; Gaps 0;

QY 112 ATTTTATTTAGGTTAAATTTTAAAGGTTTAAAGACGTACAAATCTGATGTTCAAAAT 171  
DB 311 AATTGATTAAGATTAATTAATTAACATTTGATTAATTTCTGATTAACATGTTTCCCAT 370  
QY 172 GTGACTGTGTTTAAAGAACATTTTAAATTTTAAATCTGATCAGAAATTTTAAACAAACAT 231  
DB 371 TTGATTCATTTAAACCATTAATCATCATCATCATCATCATCATCATCATCATCATCATCAT 430  
QY 232 TCTCAGTGTCTTACCCATTAATTTGATTCAGTATAGCCAGAGTGGGGGTGCTGAGA 291  
DB 431 TAAAAATAGATCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 490  
QY 292 ACTCTGAATTTTAAACGAAAGACCTCAGAAATTTCTGAAACATTTCTTAATATCA 351  
DB 491 TTGATTAAGATTAATTAATTTGATTAATTAATTAATTAATTAATTAATTAATTAATTTAT 550  
QY 352 CTCATCTAGATGAGAAATTTTGGAAAGGACCTTGAAGGCTCCAGAAAGAGTCT 411  
DB 551 CTAAATCGTTTGTGTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 610  
QY 412 CAATCAGTCTTAAAGTAACTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 471  
DB 611 CAAGGATTAACACTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 670  
QY 472 TATCTACTCTCTT 485  
DB 671 TAGATGATTAATTCAT 684

RESULT 11  
US-09-601-198-52/C  
;; Sequence 52, Application US/09601198  
;; Patent No. 6531583  
;; GENERAL INFORMATION:  
;; APPLICANT: Cassell, Gail H.  
;; APPLICANT: Chen, Ellison Y.  
;; APPLICANT: Glaes, Jennifer S.  
;; APPLICANT: Glaes, John I.  
;; APPLICANT: Heiner, Cheryl R.  
;; APPLICANT: Lefkowitz, Elliot  
;; TITLE OF INVENTION: NUCLEIC ACID PROBES AND METHOD FOR DETECTING UREAPLASMA  
;; FILE REFERENCE: UAB-13452/22  
;; CURRENT APPLICATION NUMBER: US/09/601,198  
;; PRIOR FILING DATE: 2000-12-08  
;; PRIOR APPLICATION NUMBER: 60/073,189  
;; PRIOR FILING DATE: 1998-01-30  
;; NUMBER OF SEQ ID NOS: 181  
;; SOFTWARE: Patentin Ver. 2.0  
;; SEQ ID NO 52  
;; LENGTH: 2478  
;; TYPE: DNA  
;; ORGANISM: Ureaplasma urealyticum  
US-09-601-198-52

Query Match 5.7%; Score 38.6; DB 4; Length 2478;  
Best Local Similarity 46.5%; Pred. No. 0.45;  
Matches 125; Conservative 0; Mismatches 144; Indels 0; Gaps 0;

QY 70 AATTCCAAATTCACAGCTGGGAATTAATAATTCGCCCAATTTTATTTAGGGTTAA 129  
DB 2177 AATCACCATCTTCTTAAGTCTTAATAAGATAGTCTTAATTAATTTAGCTGAAGAA 2118  
QY 130 ATTTAAGGTTTAAAGACATTAAGATCTGATGTTTCAAAATGATGATGTTTAAAGAA 189  
DB 2117 ATTAAGTAAATTTTGTGGGTCTTCAATTTGTTAAAGAAATTAATTAAGAAAGAG 2058  
QY 190 CATTTTAAATTTTATGATCAAGATTTTAAACAAATTTCTGATGCTTCAACC 249  
DB 2057 GAATTTTAATGCTTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1998  
QY 250 ATATTCATTAATGATTAAGTAAAGTGGGGGTGCTGAACTGGAATTTTAAAG 309  
DB 1997 GTGATTAAGCTTCAATGCGGTATTAATTAATTAATTAATTAATTAATTAATTAATTA 1938  
QY 310 CAGAACTCAGAAATTTCTGAAACAA 338  
DB 1937 AAAAAATACCAAAATTTCTTTAAAAA 1909

RESULT 12  
US-09-328-352-2015/C  
;; Sequence 2015, Application US/09328352  
;; Patent No. 6562958  
;; GENERAL INFORMATION:  
;; APPLICANT: Gary L. Breton et al.  
;; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
;; FILE REFERENCE: GTC99-03PA  
;; CURRENT APPLICATION NUMBER: US/09/328,352  
;; PRIOR FILING DATE: 1999-06-04  
;; NUMBER OF SEQ ID NOS: 8252  
;; SEQ ID NO 2015  
;; LENGTH: 1398  
;; TYPE: DNA  
;; ORGANISM: Acinetobacter baumannii  
US-09-328-352-2015

Query Match 5.5%; Score 37.6; DB 4; Length 1398;  
Best Local Similarity 55.3%; Pred. No. 0.67;  
Matches 73; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY 100 AATGCTCCATGATTTTATTAAGGTTAAATTTAAGGTTTAAAGCAGTACAGATCTG 159  
DB 355 AATCTTCTTAATTTCAAAATTTGATTAATTAATTAATTAATTAATTAATTAATTAATTA 296  
QY 160 AATGTCACAAATGCTGCTGTTAAGAAACATTTTAAATCTGATCAGAAATTTTA 219  
DB 295 AAGCAAAAATTAAGCAAAATTTCCGATTAATTTTCAATGCTTTAGTATCTATGT 236  
QY 220 AAACAAAACAT 231  
DB 235 TACTAATGACAT 224

RESULT 13  
US-09-328-352-1803  
;; Sequence 1803, Application US/09328352  
;; Patent No. 6562958  
;; GENERAL INFORMATION:  
;; APPLICANT: Gary L. Breton et al.  
;; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
;; FILE REFERENCE: GTC99-03PA  
;; CURRENT APPLICATION NUMBER: US/09/328,352  
;; PRIOR FILING DATE: 1999-06-04  
;; NUMBER OF SEQ ID NOS: 8252  
;; SEQ ID NO 1803  
;; LENGTH: 270  
;; TYPE: DNA  
;; ORGANISM: Acinetobacter baumannii  
US-09-328-352-1803

Query Match	5.5%	Score 37.4	DB 4	Length 270
Best Local Similarity	55.9%	Pred. No. 0.39		
Matches 71	Conservative 0	Mismatches 56	Indels 0	Gaps 0

Qy	109	CAGATTTTAA	TTTAAAGGTTAA	ATTAAAGGTTTAA	GACATCAAGATCTGATTCACA	168
Db	13	CAAAATTTTA	TTTAAATTCATTTAA	TTTTTTTAA	TTTAAATGATGACTTAATTATATA	72
Qy	169	AATGTGACTGT	TTTAAAGAAACATTTTAAATCTTGATCA	CAGATTTTAAACAAAAA	228	
Db	73	TTTGAATTTGAT	TTTGATTAATTAATTAACTTTTAATATATTAATTTAAAAATAT	132		
Qy	229	CATTCTC	235			
Db	133	AATTATC	139			

RESULT 14  
US-09-248-796A-4289/C  
; Sequence 4289, Application US/09248796A

```

; GENERAL INFORMATION:
; APPLICANT: Keith Weinstein et al
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 4289
; LENGTH: 927
; TYPE: DNA
; ORGANISM: Candida albicans
; US-09-248-796A-4289

```

Query Match	5.5%	Score 37.2;	DB 4;	Length 927;
Best Local Similarity	50.6%;	Pred. No. 0.74;		
Matches	90;	Mismatches	88;	Indels 0;
				Gaps 0;

QY 89 TTGGAAATATAAAATCTGCCTCAGATTTTATTTAGGGTTAAATTTAGGGTTTAAACAC 148

Db 372 TTGGCTTTATGTCTTCCCAAGATTTTTTTTTCACCAATGTTTGAAGATGATTA 313

QY 149 TACAAGATCTGATGTTTACAAATGTAAGTGTGTTTAAAGAACATTTTAAATCTTGATC 208

Db 312 CCCAAATTCGTCTGTGACATATCGCTCTGAGAGAAAACATGACAACTTCAAAATC 253

QY 209 ACAAGATTTTAAACAAAACATTTCTCAGTGCCTTACCCATATTCGATTTAGAT 266

Db 252 CCAGAGATATCTTAAATTTCCATTCATTCATCTTTGATGTTGATTTTATTCACATT 195

```

RESULT 15
US-09-134-001C-2435
; Sequence 2435, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674

```

```

; SEQ ID NO 2435
;
; LENGTH: 357
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-2435

```

Query Match	5.4%	Score 36.8	DB 3	Length 357
Best Local Similarity	53.5%	Pred. No. 0.64		
Matches 77; Conservative	0	Mismatches 67	Indels 0	Gaps 0

OY	8	GCTGGAAATAAATACCTCCCAAGATTAAATGAGGAATTAATGAGGTGTTTGC	17
Db	118	GGTTGAAAGTTTGAAATGCAATTAAGAAGCTCCTTTGGAGTTAAGCGTAGGAATGGAT	177
OY	147	AGTACAAGATCTGATNTTTCACAAATGTGACTGTGTTTAAGAAACATTTTTAAATCTTGA	266
Db	178	AGTTTAAAACCTTAAGAGGACGGAATCGCTGATTATAAGAGAAATGCGACAACTGAGA	237
OY	207	TCAAGAATTTTAAATCAAATAACA	230
Db	238	TTACAAAGATTGAACGTAAAGAA	261

Search completed: February 8, 2005, 22:26:29  
Job time : 185.767 secs



QY 109 CAGATTTTATTAGGTTAAATTTAAGGTTTAAAGACATGACATGATGTTTCA 168  
DB 69991 -----TCATAGTTGGTGAATTAAGGTTTAAACAGTATGATGATGATGTTTCA 70041  
QY 169 AATGTACTGTGTGTTAAGAAACATTTTAAATCTTGATCAAGATTTTAAAAACAAA 228  
DB 70042 TATGTCTCTGGGTTGAAGAAACCTTCAGGAGCTT-----GTTTAAAAAGTA 70090  
QY 229 CATTCCTCAGTGTGCTTACCCCATATCTGATTCAGATAGCCAGATGGGGGGTGTGCG 288  
DB 70091 TATTCCTCAATGCGGCTTACCAAAA-----ATTCTGATTTGGTACACTGGGGGGGGCC 70145  
QY 289 AGAAGCTGAAATTTTAAACAGCA-----AGAAGCTGAGAAATCTTGAAACAACTACTT 344  
DB 70146 AGGACTCTGATTTTAAAGCAACCCAGAGATGCTGTTGAAGCTGTAGCTGTAAAT 70205  
QY 345 ATTAATCACTCATCTCTAGATGGA--AAATTTTGAAGGAGACTTTGAAAGGCTTCAGA 402  
DB 70206 ATCAACACCCATCTCTAGATGAGAGAGCTTTTGAAGGAGCCCTTGAAGGCTTCAGA 70265  
QY 403 AAAAGCTCATCATCTCTT---AGTACTACAAAAATGCCGTTTCTTAAACCA 459  
DB 70266 GAAAGGCTTAACACACTTGTGACCAATATTAAGAGATGCCGTTTGTCTAAACCA 70325  
QY 460 ATTCTCAGAGATATCACTCTCTTCTGCTCAGAAAAACCCGGGCTATCATCTTCA 519  
DB 70326 ATTCTCTCAAGATTTCAATCTCTTCTGCTCAGCAATATGCTGTATACCTCTCA 70385  
QY 520 --GAGTGAATTTCACTGTTTGAAGATTTCTTCTGAACTATTTATTTGT 567  
DB 70386 GGGGGTGAAGATTTTGTGTTAGGAATTCACCTTTTGAAGCCACATCTCTGT 70435

RESULT 2  
US-10-027-632-287050/c  
; Sequence 287050, Application US/10027632  
; Publication No. US20020198371A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027, 632  
; PRIOR FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218, 006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198, 676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193, 483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185, 218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167, 363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156, 358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146, 002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 287050  
; LENGTH: 600  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-287050

Query Match 11.8%; Score 80.4; DB 13; Length 600;  
Best Local Similarity 70.4%; Pred. No. 1.3e-10;  
Matches 145; Conservative 1; Mismatches 37; Indels 23; Gaps 2;  
QY 1 ATGGGAGCTGATTTTGTGCTGCTTCTGGAGAGAGCTTCAAGATGCTGTGAGT 60  
DB 256 ATGGGAGCTGATTTTGTGCTGCTTCTGGAGAGAGCTTCAAGATGCTGTGAGT 197

QY 61 AAAATTTT-----CAATTTCAATTTCAAGCTTGAATTAATGTCGCT 108  
DB 196 AAAACACCCCTTGATTAAGTCAAGTGTGATTTCAAACTGATTAATGTCGCT- 138  
QY 109 CAGATTTTATTAGGTTAAATTTAAGGTTTAAAGACATGACATGATGTTTCA 168  
DB 137 -----CATAGTTGGTGAATTTAAGGTTTAAACAGTATGATCAGATGTCCTCA 88  
QY 169 AATGTACTGTGTGTTAAGAAACATTT 194  
DB 87 TATGTCTCTGGGTTGAAGAAACACTT 62

RESULT 3  
US-10-027-632-287050/c  
; Sequence 287050, Application US/10027632  
; Publication No. US20030204075A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027, 632  
; PRIOR FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218, 006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198, 676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193, 483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185, 218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167, 363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156, 358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146, 002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 287050  
; LENGTH: 600  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-287050

Query Match 11.8%; Score 80.4; DB 15; Length 600;  
Best Local Similarity 70.4%; Pred. No. 1.3e-10;  
Matches 145; Conservative 1; Mismatches 37; Indels 23; Gaps 2;  
QY 1 ATGGGAGCTGATTTTGTGCTGCTTCTGGAGAGAGCTTCAAGATGCTGTGAGT 60  
DB 256 ATGGGAGCTGATTTTGTGCTGCTTCTGGAGAGAGCTTCAAGATGCTGTGAGT 197  
QY 61 AAAATTTT-----CAATTTCAATTTCAAGCTTGAATTAATGTCGCT 108  
DB 196 AAAACACCCCTTGATTAAGTCAAGTGTGATTTCAAACTGATTAATGTCGCT- 138  
QY 109 CAGATTTTATTAGGTTAAATTTAAGGTTTAAAGACATGACATGATGTTTCA 168  
DB 137 -----CATAGTTGGTGAATTTAAGGTTTAAACAGTATGATCAGATGTCCTCA 88  
QY 169 AATGTACTGTGTGTTAAGAAACATTT 194  
DB 87 TATGTCTCTGGGTTGAAGAAACACTT 62

RESULT 4  
US-10-027-632-287049/c  
; Sequence 287049, Application US/10027632  
; Publication No. US20020198371A1  
; GENERAL INFORMATION:



```
/ APPLICANT: Wang, David G.
/ TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
/ FILE REFERENCE: 109827.129
/ CURRENT APPLICATION NUMBER: US/10/027,632
/ PRIOR FILING DATE: 2002-04-30
/ PRIOR APPLICATION NUMBER: US 60/218,006
/ PRIOR FILING DATE: 2000-07-12
/ PRIOR APPLICATION NUMBER: US 60/198,676
/ PRIOR FILING DATE: 2000-04-20
/ PRIOR APPLICATION NUMBER: US 60/193,483
/ PRIOR FILING DATE: 2000-03-29
/ PRIOR APPLICATION NUMBER: US 60/185,218
/ PRIOR FILING DATE: 1999-11-23
/ PRIOR APPLICATION NUMBER: US 60/167,363
/ PRIOR FILING DATE: 1999-09-28
/ PRIOR APPLICATION NUMBER: US 60/146,002
/ NUMBER OF SEQ ID NOS: 325720
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 287049
/ LENGTH: 600
/ TYPE: DNA
/ ORGANISM: Human
US-10-027-632-287049

Query Match      11.6%; Score 79.2; DB 13; Length 600;
Best Local Similarity 66.2%; Pred. No. 2.6e-10;
Matches 129; Conservative 0; Mismatches 65; Indels 1; Gaps 1;

QY      1  ATGGGACCTGATTTTGTGCTGCTGCTGCGAGACGCTTCAGTAGCTGAGT 60
DB      256  ATGGGACCTGATTTTATTTGCTGCTGCGAGACGCTTCAGTAGCTGAGT 197
QY      61  AAAATTTCCAAATTTCCAAATTTCCAAAGCTTGAATAATTAATCTGCTCAGATTTTAT- 119
DB      196  AAAACACCCCTTGCAATGACAGTGCATTTCCAAATTTGACATTAATAATCTGCTC 137
QY      120  TTAGGCTTAATTTAAGGTTTAAAGACAGTACAGATCTGATGTTCAAAATGTGACTGT 179
DB      136  ATAGTTGCTGAATTTAGGGTTTAAACAGATGATGATGTCTTCAATATGTCTCTGG 77
QY      180  GTTTAGAAACATTT 194
DB      76  GTTGAAGAAACACTT 62

RESULT 5
US-10-027-632-287049/C
/ Sequence 287049, Application US/10027632
/ Publication No. US20030204075A9
/ GENERAL INFORMATION:
/ APPLICANT: Wang, David G.
/ TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
/ FILE REFERENCE: 109827.129
/ CURRENT APPLICATION NUMBER: US/10/027,632
/ PRIOR FILING DATE: 2002-04-30
/ PRIOR APPLICATION NUMBER: US 60/218,006
/ PRIOR FILING DATE: 2000-07-12
/ PRIOR APPLICATION NUMBER: US 60/198,676
/ PRIOR FILING DATE: 2000-04-20
/ PRIOR APPLICATION NUMBER: US 60/193,483
/ PRIOR FILING DATE: 2000-03-29
/ PRIOR APPLICATION NUMBER: US 60/185,218
/ PRIOR FILING DATE: 2000-02-24
/ PRIOR APPLICATION NUMBER: US 60/167,363
/ PRIOR FILING DATE: 1999-11-23
/ PRIOR APPLICATION NUMBER: US 60/156,358
/ PRIOR FILING DATE: 1999-09-28
/ PRIOR APPLICATION NUMBER: US 60/146,002
```

```
/ PRIOR FILING DATE: 1999-08-09
/ NUMBER OF SEQ ID NOS: 325720
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 287049
/ LENGTH: 600
/ TYPE: DNA
/ ORGANISM: Human
US-10-027-632-287049

Query Match      11.6%; Score 79.2; DB 15; Length 600;
Best Local Similarity 66.2%; Pred. No. 2.6e-10;
Matches 129; Conservative 0; Mismatches 65; Indels 1; Gaps 1;

QY      1  ATGGGACCTGATTTTGTGCTGCTGCTGCGAGACGCTTCAGTAGCTGAGT 60
DB      256  ATGGGACCTGATTTTATTTGCTGCTGCGAGACGCTTCAGTAGCTGAGT 197
QY      61  AAAATTTCCAAATTTCCAAATTTCCAAAGCTTGAATAATTAATCTGCTCAGATTTTAT- 119
DB      196  AAAACACCCCTTGCAATGACAGTGCATTTCCAAATTTGACATTAATAATCTGCTC 137
QY      120  TTAGGCTTAATTTAAGGTTTAAAGACAGTACAGATCTGATGTTCAAAATGTGACTGT 179
DB      136  ATAGTTGCTGAATTTAGGGTTTAAACAGATGATGATGTCTTCAATATGTCTCTGG 77
QY      180  GTTTAGAAACATTT 194
DB      76  GTTGAAGAAACACTT 62

RESULT 6
US-10-104-774-3
/ Sequence 3, Application US/10104774
/ Publication No. US20020164630A1
/ GENERAL INFORMATION:
/ APPLICANT: Perkin-Elmer Corporation, Applied Biosystems Division
/ TITLE OF INVENTION: DETERMINATION OF GENOTYPE OF AMPLIFICATION PRODUCTS AT MULTIPLE
/ FILE REFERENCE: 16842-782
/ CURRENT APPLICATION NUMBER: US/10/104,774
/ PRIOR FILING DATE: 2002-03-21
/ PRIOR APPLICATION NUMBER: US 09/018,595
/ PRIOR FILING DATE: 1998-02-04
/ NUMBER OF SEQ ID NOS: 11
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 3
/ LENGTH: 500
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-10-104-774-3

Query Match      6.6%; Score 44.8; DB 13; Length 500;
Best Local Similarity 87.5%; Pred. No. 0.25;
Matches 49; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY      1  ATGGGACCTGATTTTGTGCTGCTGCTGCGAGACGCTTCAGTAGCTGCTGT 56
DB      69  ATGGGACCTGATTTTATTTGCTGCTGCGAGACGCTTCAGTAGCTGCTGT 124

RESULT 7
US-10-455-150-3
/ Sequence 3, Application US/10455150
/ Publication No. US20040053302A1
/ GENERAL INFORMATION:
/ APPLICANT: Perkin-Elmer Corporation, Applied Biosystems Division
/ TITLE OF INVENTION: FLUORESCENCE GENOTYPING AT MULTIPLE ALLELIC
/ NUMBER OF SEQUENCES: 11
/ CORRESPONDENCE ADDRESSES:
/ ADDRESSEE: David J. Weitz,
/ Wilson Sonsini Goodrich & Rosati
```

STREET: 650 Page Mill Road  
CITY: Palo Alto  
STATE: California  
COUNTRY: USA  
ZIP: 94304-1050  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch diskette  
COMPUTER: IBM compatible  
OPERATING SYSTEM: Microsoft Windows 95  
SOFTWARE: Wordperfect for windows 6.0,  
ASCII (DOS) TEXT format  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/455,150  
FILING DATE: 04-Jun-2003  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/324,709A  
FILING DATE: 03-June-1999  
ATTORNEY/AGENT INFORMATION:  
NAME: David J. Weitz  
REGISTRATION NUMBER: 38,362  
REFERENCE/DOCKET NUMBER: 16842-758  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 493-9300  
TELEFAX: (650) 493-6811  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 500 nucleotides  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 3:  
US-10-455-150-3

Query Match 6.6%; Score 44.8; DB 16; Length 500;  
Best Local Similarity 87.5%; Pred. No. 0.25;  
Matches 49; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 ATGGGACCTGGATTGTTGCTGCTCTCTGGAGACGCTTCAGTATGCTGT 56  
DB 69 ATGGGACCTGGATTGTTGCTGCTCTCTGGAGACGCTTCAGTATGCTGT 124

RESULT 8  
US-10-104-774-1  
Sequence 1, Application US/10104774  
Publication No. US20020164630A1  
GENERAL INFORMATION:  
APPLICANT: Perkin-Elmer Corporation, Applied Biosystems Division  
TITLE OF INVENTION: DETERMINATION OF GENOTYPE OF AMPLIFICATION PRODUCTS AT MULTIPLE A  
TITLE OF INVENTION: SITES  
FILE REFERENCE: 16842-782  
CURRENT APPLICATION NUMBER: US/10/104,774  
CURRENT FILING DATE: 2002-03-21  
PRIOR APPLICATION NUMBER: US 09/018,595  
PRIOR FILING DATE: 1998-02-04  
NUMBER OF SEQ ID NOS: 11  
SOFTWARE: Patentin version 3.1  
SEQ ID NO 1  
LENGTH: 793  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-104-774-1

Query Match 6.6%; Score 44.8; DB 13; Length 793;  
Best Local Similarity 87.5%; Pred. No. 0.31;  
Matches 49; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 ATGGGACCTGGATTGTTGCTGCTCTCTGGAGACGCTTCAGTATGCTGT 56  
DB 69 ATGGGACCTGGATTGTTGCTGCTCTCTGGAGACGCTTCAGTATGCTGT 124

RESULT 9  
US-10-455-150-1  
Sequence 1, Application US/10455150  
Publication No. US20040053302A1  
GENERAL INFORMATION:  
APPLICANT: Perkin-Elmer Corporation,  
Applied Biosystems Division  
TITLE OF INVENTION: FLUORESCENCE GENOTYPING AT MULTIPLE ALLELIC  
SITES  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSER: David J. Weitz,  
Wlason Sonstini Goodrich & Rosati  
STREET: 650 Page Mill Road  
CITY: Palo Alto  
STATE: California  
COUNTRY: USA  
ZIP: 94304-1050  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch diskette  
COMPUTER: IBM compatible  
OPERATING SYSTEM: Microsoft Windows 95  
SOFTWARE: Wordperfect for windows 6.0,  
ASCII (DOS) TEXT format  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/455,150  
FILING DATE: 04-Jun-2003  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/324,709A  
FILING DATE: 03-June-1999  
ATTORNEY/AGENT INFORMATION:  
NAME: David J. Weitz  
REGISTRATION NUMBER: 38,362  
REFERENCE/DOCKET NUMBER: 16842-758  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 493-9300  
TELEFAX: (650) 493-6811  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 793 nucleotides  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-10-455-150-1

Query Match 6.6%; Score 44.8; DB 16; Length 793;  
Best Local Similarity 87.5%; Pred. No. 0.31;  
Matches 49; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 ATGGGACCTGGATTGTTGCTGCTCTCTGGAGACGCTTCAGTATGCTGT 56  
DB 69 ATGGGACCTGGATTGTTGCTGCTCTCTGGAGACGCTTCAGTATGCTGT 124

RESULT 10  
US-10-755-889-605  
Sequence 605, Application US/10755889  
Publication No. US20040171823A1  
GENERAL INFORMATION:  
APPLICANT: Bristol-Myers Squibb Company  
TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ASSOCIATED WITH THE NF-KB  
PATHWAY  
FILE REFERENCE: D0284 NP  
CURRENT APPLICATION NUMBER: US/10/755,889  
CURRENT FILING DATE: 2004-01-13  
PRIOR APPLICATION NUMBER: U.S. 60/440,068  
PRIOR FILING DATE: 2003-01-14  
PRIOR APPLICATION NUMBER: U.S. 60/469,757  
PRIOR FILING DATE: 2003-05-12  
NUMBER OF SEQ ID NOS: 823  
SOFTWARE: Patentin version 3.2

SEQ ID NO 605  
LENGTH: 793  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-755-889-605

Query Match 6.4%; Score 44.8; DB 17; Length 793;  
Best Local Similarity 87.5%; Pred. No. 0.31;  
Matches 49; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 ATGGGACCTGATTTGTTGCTGCTGCTGAGACACCTTCAGTATGCTGT 56  
Db 69 ATGGGACCTGATTTATTTGCTGCTGCTGAGACACCTTTTGCATGCTCT 124

RESULT 11  
US-10-104-774-2  
Sequence 2, Application US/10104774  
Publication No. US20020164630A1  
GENERAL INFORMATION:

APPLICANT: Perkin-Elmer Corporation, Applied Biosystems Division  
TITLE OF INVENTION: DETERMINATION OF GENOTYPE OF AMPLIFICATION PRODUCTS AT MULTIPLE A  
FILE REFERENCE: 16842-782  
CURRENT APPLICATION NUMBER: US/10/104,774  
CURRENT FILING DATE: 2002-03-21  
PRIOR FILING DATE: 1998-02-04  
NUMBER OF SEQ ID NOS: 11  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 2  
LENGTH: 802  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-104-774-2

Query Match 6.4%; Score 43.2; DB 13; Length 802;  
Best Local Similarity 85.7%; Pred. No. 0.82;  
Matches 48; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 1 ATGGGACCTGATTTGTTGCTGCTGCTGAGACACCTTCAGTATGCTGT 56  
Db 69 ATGGGACCTGATTTGTTGCTGCTGCTGAGACACCTTTTGCATGCTCT 124

RESULT 12  
US-10-455-150-2  
Sequence 2, Application US/10455150  
Publication No. US20040053302A1  
GENERAL INFORMATION:

APPLICANT: Perkin-Elmer Corporation,  
TITLE OF INVENTION: FLUORESCENCE GENOTYPING AT MULTIPLE ALLELIC  
SITES  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSES: David J. Weitz,  
Wilson Sonsini Goodrich & Rosati  
STREET: 650 Page Mill Road  
CITY: Palo Alto  
STATE: California  
COUNTRY: USA  
ZIP: 94304-1050  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch diskette  
COMPUTER: IBM compatible  
OPERATING SYSTEM: Microsoft Windows 95  
SOFTWARE: Wordperfect for windows 6.0,  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/455,150  
FILING DATE: 04-Jun-2003  
CLASSIFICATION: <Unknown>

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/324,709A  
FILING DATE: 03-June-1999  
ATTORNEY/AGENT INFORMATION:  
NAME: David J. Weitz  
REGISTRATION NUMBER: 38,362  
REFERENCE/DOCKET NUMBER: 16842-758  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 493-9300  
TELEFAX: (650) 493-6811  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 802 nucleotides  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
US-10-455-150-2

Query Match 6.4%; Score 43.2; DB 16; Length 802;  
Best Local Similarity 85.7%; Pred. No. 0.82;  
Matches 48; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 1 ATGGGACCTGATTTGTTGCTGCTGCTGAGACACCTTCAGTATGCTGT 56  
Db 69 ATGGGACCTGATTTGTTGCTGCTGCTGAGACACCTTTTGCATGCTCT 124

RESULT 13  
US-10-723-860-5797  
Sequence 5797, Application US/10723860  
Publication No. US20040253606A1  
GENERAL INFORMATION:

APPLICANT: Aziz, Natasha  
APPLICANT: Ginsburg, Wendy M.  
TITLE OF INVENTION: Methods for Diagnosis of Soft Tissue Sarcoma, Compositions &  
FILE REFERENCE: 05882.0193.NPUS01  
CURRENT FILING DATE: 2003-11-26  
PRIOR APPLICATION NUMBER: 60/429,739  
PRIOR FILING DATE: 2002-11-26  
NUMBER OF SEQ ID NOS: 8393  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 5797  
LENGTH: 852  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-723-860-5797

Query Match 6.4%; Score 43.2; DB 18; Length 852;  
Best Local Similarity 85.7%; Pred. No. 0.84;  
Matches 48; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 1 ATGGGACCTGATTTGTTGCTGCTGCTGAGACACCTTCAGTATGCTGT 56  
Db 69 ATGGGACCTGATTTGTTGCTGCTGCTGAGACACCTTTTGCATGCTCT 124

RESULT 14  
US-10-450-826-53  
Sequence 53, Application US/10450826  
Publication No. US20040101818A1  
GENERAL INFORMATION:

APPLICANT: JI, Darren  
APPLICANT: Cook, Jonathan S.  
APPLICANT: Jalewal, Neelam  
APPLICANT: Bistein, Richard  
APPLICANT: Houghton, Adam  
APPLICANT: Weitz, Lawrence  
TITLE OF INVENTION: Gene Expression Profiles Associated with Osteoblast Differentiat

FILE REFERENCE: 044921-5039-WO  
CURRENT APPLICATION NUMBER: US/10/450,826  
CURRENT FILING DATE: 2003-06-18  
PRIOR APPLICATION NUMBER: US 60/255,882  
PRIOR FILING DATE: 2000-12-18  
PRIOR APPLICATION NUMBER: US 60/285,691  
PRIOR FILING DATE: 2001-04-24  
NUMBER OF SEQ ID NOS: 149  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 53  
LENGTH: 178896  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: Genbank Accession No. AC008041  
US-10-450-826-53

Query Match 6.3%; Score 43; DB 17; Length 178896;  
Best Local Similarity 46.5%; Pred. No. 14;  
Matches 139; Conservative 0; Mismatches 160; Indels 0; Gaps 0;

QY 165 CAGAAATGACGCTGTTTAAAGAAATTTTAAATCTTGATCACAAGATTTTAAACA 224  
DB 73412 CAAAAATGCAAAATGATCAATCAAGTTAAAGCTTCTGCACAGCCAGATACAA 73471  
QY 225 AAAACATCTCAGTTCCTTACCCATATTCGATTCAGTAAGCCAGAGTGGGGGTGT 284  
DB 73472 CCAACAAAGTAGAGACCAACCAAGAAATATCTGTAACTACCCATCTG 73531  
QY 285 GCTGAGAACTGTAATTTTAAAGCAAGAACTCAGAAATCTTGAAACATTAATT 344  
DB 73532 ACAGAGATTATTAACCAATATATAGAGCTCAACAACTCTATAGGAAAAATCTA 73591  
QY 345 ATAATCACTATCTGATGAGAAATTTTGAAGGACCTTGAAAGGCTCCAGAAA 404  
DB 73592 ATATCTGATTTAAAAATGGGCAAAAGATTTGAAATGACATTTCTCAAAAGAACATA 73651  
QY 405 AAGTGCTCAATCAGTCTTTAAAGTACTCAAAAATGCAAGTTTCTTAAACCAATTT 463  
DB 73652 CAGAGGGCAAAACAGCATTTGATTAATCAGATTAATGAATAATCAACAACTATATCATAT 73710

RESULT 15  
US-10-417-375-66

Sequence 66, Application US/10417375  
Publication No. US20040219528A1  
GENERAL INFORMATION:  
APPLICANT: David W. Morris  
APPLICANT: Marc Malandro  
TITLE OF INVENTION: Novel Therapeutic Targets in Cancer  
FILE REFERENCE: 529452001600  
CURRENT APPLICATION NUMBER: US/10/417,375  
CURRENT FILING DATE: 2003-04-15  
NUMBER OF SEQ ID NOS: 176  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 66  
LENGTH: 252907  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-417-375-66

Query Match 6.2%; Score 42.4; DB 18; Length 252907;  
Best Local Similarity 52.2%; Pred. No. 24;  
Matches 94; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

QY 76 CAATTCACAGCTTGAAATTAATCTGCTCAGATTTTATTTAGGGTTAAATTTAA 135  
DB 177651 CACATGATATCTTTTAAAAAACAATGATCATCTTTTATTAATTTGATTTT 177710  
QY 136 GGGTTTAAGACATCAAGATTCGATGTTCAAAATGATGCTGTTTAAAGAAATTT 195  
DB 177711 AAATTTAAAGATATATGAAATATTTTCCATATTTGATGATATATATATATAG 177770

QY 196 TAAATCTGATCACAAGATTTTAAACAAACATTCAGTGCCTGACCCATATTC 255  
DB 177771 TATAATATGATGTTATGATATATATATATATATTTTGAAGATGAGCTCACTTGT 177830  
Search completed: February 9, 2005, 06:51:29  
Job time : 925.298 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 8, 2005, 12:58:25 ; Search time 8099.44 Seconds  
(without alignments)  
3059.349 Million cell updates/sec

Title: US-10-754-437-23

Perfect score: 680  
1 atgsggagactgagatttgc.....gctaactgagagacagacac 680

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 32822875 seqs, 1821965908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_est9:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	90.6	13.3	516	8	AQ003910	AQ003910 CIT-HSP-2
2	49.6	7.3	429	6	CB473334	CB473334 sm68_E03.
3	46.6	6.9	303	7	W33932	W33932 mb54d02.r1
4	46.6	6.9	430	7	W40649	W40649 mb43h12.r1
5	46.6	6.9	536	7	W36345	W36345 mb72c12.r1
6	46.6	6.9	585	9	AY419441	AY419441 Mus muscu
7	46.6	6.9	645	6	CD773419	CD773419 AGENCOURT
8	46.6	6.9	707	6	CB056709	CB056709 NISC_j119
9	46.6	6.9	812	3	AK029358	AK029358 Mus muscu
10	46.6	6.9	843	6	CB588212	CB588212 AGENCOURT
11	46.6	6.9	859	6	CB588525	CB588525 AGENCOURT
12	46.6	6.9	865	6	CB589177	CB589177 AGENCOURT
13	46.6	6.9	886	6	CB587051	CB587051 AGENCOURT
14	46.6	6.9	889	6	CB589251	CB589251 AGENCOURT
15	46.6	6.9	891	6	CB574837	CB574837 AGENCOURT
16	46.6	6.9	913	6	CB590451	CB590451 AGENCOURT
17	46.6	6.9	919	6	CB587332	CB587332 AGENCOURT
18	46.6	6.9	928	6	CB590111	CB590111 AGENCOURT
19	46.6	6.9	945	2	BB614068	BB614068 BB614068
20	46.6	6.9	990	7	W12906	W12906 ma89b03.r1
21	46.6	6.9	1006	7	W29475	W29475 mb99f11.r1
22	46.6	6.9	1020	7	W08102	W08102 mb40g08.r1
23	46.6	6.9	1101	9	CNS0039G	AL063921 Drosophi1
24	45.6	6.7	293	7	R47143	R47143 Y79 Rat inc

25	45.6	6.7	556	7	R46934	R46934 Y169 Rat in
26	45	6.6	217	7	R46913	R46913 Y140 Rat in
27	45	6.6	223	2	BB571643	BB571643 BB571643
28	45	6.6	395	7	R46955	R46955 Y195 Rat in
29	45	6.6	449	7	R47024	R47024 Y350 Rat in
30	45	6.6	1031	6	CB575508	CB575508 AGENCOURT
31	44.8	6.6	529	9	CR342574	CR342574 Medicago
32	44.8	6.6	570	9	AY419439	AY419439 Homo sapi
33	44.8	6.6	570	9	AY419440	AY419440 Pan trogl
34	44.6	6.6	303	7	R46903	R46903 Y124 Rat in
35	44.6	6.6	318	7	R47078	R47078 Y534 Rat in
36	44.6	6.6	467	7	R46947	R46947 Y186 Rat in
37	44.4	6.5	987	9	CNS014PQ	AL106456 Drosophi1
38	44	6.5	373	7	R47135	R47135 Y722 Rat in
39	44	6.5	395	7	R47030	R47030 Y359 Rat in
40	43.8	6.4	1101	9	CNS00386	AL064084 Drosophi1
41	43.8	6.4	1313	7	CK997149	CK997149 ip15C02.b
42	43.4	6.4	257	7	R47100	R47100 Y601 Rat in
43	43.2	6.4	727	4	BH705290	BH705290 BOHVM06TF
44	43.2	6.4	783	4	BG198114	BG198114 RST17499
45	43.2	6.4	1101	9	CNS0039G	AL063921 Drosophi1

#### ALIGNMENTS

RESULT 1  
LOCUS AQ003910/c 516 bp DNA linear GSS 26-JUN-1998  
DEFINITION CIT-HSP-2289L4.TF CIT-HSP Homo sapiens genomic clone 2289L4,  
Genomic survey sequence.

ACCESSION AQ003910  
VERSION AQ003910.1 GI:3081561  
KEYWORDS GSS.

SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 516)  
AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Linher,K.,  
Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,  
Simon,M. and Venter,J.C.  
Use of a random BAC End Sequence Database for Sequence-Ready Map

TITLE Building (1998)  
JOURNAL Unpublished (1998)  
COMMENT Other GSSs: CIT-HSP-2289L4.TR  
Contact: Mark Adams  
Department of Bukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: mdadams@tigr.org  
Clones are available from Research Genetics (info@resgen.com). BAC  
end search page:  
http://www.tigr.org/cdb/hungen/bac\_end\_search/bac\_end\_search.html  
Seq primer: M13-21;  
Class: BAC ends.

FEATURES  
source Location/Qualifiers  
1..516  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="GDB:7150375"  
/db\_xref="taxon:9606"  
/clone="2289L4"  
/sex="Male"  
/cell\_type="Sperm"  
/clone\_lib="CIT-HSP"  
/note="Vector: pBeloBAC11, Site\_1: HindIII, Site\_2:  
HindIII"

ORIGIN  
Query Match 13.3%; Score 90.6; DB 8; Length 516;

Best Local Similarity 69.0%; Pred. No. 7,4e-12;  
Matches 169; Conservative 0; Mismatches 69; Indels 7; Gaps 3;

QY 330 TGAACAATTACTTATTAATCACTCATCTCTAGATG--GAAATTTTGGAGGACCTT 387  
DB 493 TGTACTTGTATTAATCAACCCATCTATAGATGAGAAAGGCTTTTGGAGGACCTT 434  
QY 388 TGAAGGCTCCAGAGAAAGTGTCAATGAGTCT--TAAGTACACAAAATGACAGT 444  
DB 433 TGAAGGCTCCAGAGAAAGTGTCAACGCTTGGACAAATATGACAGATGCCAGT 374  
QY 445 TTTCTTAAACCAATTTCTCAAGATATCTAATCTCTTCTGCTCAGAAAACCCCG 504  
DB 373 TTTGTCAAAACCAATCTCTCAAGATTCAAATCTCTCTGCGCTCCACATATGCG 314  
QY 505 GCGTATCATCTTCTCA--GAGTAAAGATTTCAGTGTAGAAATTTCTCTGAACTATTA 562  
DB 313 TGTCTTAAACCCCTCAGGGGCTAAGATTTTGTGTAGAAATTCATTTTGGACCCCAT 254  
QY 563 TTTGT 567  
DB 253 CTTGT 249

RESULT 2  
LOCUS CB473334 429 bp mRNA linear EST 26-MAR-2003  
DEFINITION n68\_E03.f sn Sus scrofa CDNA 5', mRNA sequence.  
ACCESSION CB473334  
VERSION CB473334.1 GI:29279720  
KEYWORDS EST.  
SOURCE Sus scrofa (pig)  
ORGANISM Sus scrofa

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.  
AUTHORS 1 (bases 1 to 429)  
TITLE Neilan,J.G., Kutish,G.F., Lu,Z., Zaak,A. and Rock,D.L.  
non-infected porcine macrophage CDNA libraries  
JOURNAL Unpublished (2003)  
COMMENT Contact: Neilan JG  
Plum Island Animal Disease Center  
US Department of Agriculture, Agricultural Research Service  
PO Box 848, Greenport, NY 11944-848, USA  
Tel: 631 323 3133  
Fax: 631 323 3044  
Email: jneilan@piadc.ars.usda.gov

Single pass sequencing. Bases called with phred v0.020425.c and  
trimmed with the aid of the trim.alg option. Vector identified by  
cross match v0.990329 and Lucy v1.17p.  
Seq primer: M13 Forward.  
FEATURES  
Location/Qualifiers  
1..429  
/organism="Sus scrofa"  
/mol\_type="mRNA"  
/db\_xref="taxon:9823"  
/tissue\_type="lymphoid"  
/cell\_type="macrophage"  
/lab\_host="DH10B"  
/clone\_lib="sn"  
/note="Vector: pSPORT1; Site 1: NotI; Site 2: SalI;  
Library made from pools of polyA selected RNA, isolated at  
different times post-infection (0 to 16 hrs) from African  
swine fever virus (isolate Pretoriuskop/96/4) infected  
swine macrophages. Macrophages were derived from  
peripheral blood mononuclear cells cultured for 48 hrs on  
plastic in the presence of 30% FCS supernatant."

## ORIGIN

Query Match 7.3%; Score 49.6; DB 6; Length 429;  
Best Local Similarity 92.9%; Pred. No. 0.13;  
Matches 52; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ATGGGACCTGGATTTTGTGCTGCTCTCTGGAGACAGCTTCAGTATGCTGT 56  
DB 368 ATGGGACCTGGATTTTGTGCTGCTCTCTGGAGACAGCTTCATATGCTCT 313

RESULT 3  
LOCUS W33932 303 bp mRNA linear EST 11-SEP-1996  
DEFINITION mb54d02.r1 Soares mouse p3NMFI9.5 Mus musculus CDNA clone  
IMAGE:333219.5 similar to gb:U10095 Mouse amelogenin (MOUSE);  
mRNA sequence.  
ACCESSION W33932  
VERSION W33932.1 GI:1315837  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
AUTHORS 1 (bases 1 to 303)  
Marra,M., Hillier,L., Allen,M., Bowler,M., Dietrich,N., Dubuque,T.,  
Giesel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,K. and  
Waterston,R.  
The WashU-HMI Mouse EST Project  
Unpublished (1996)  
CONTACT: Marra M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.edu

This clone is available royalty-free through LML; contact the  
IMGS Consortium (info@image.llnl.gov) for further information.  
MGI:214619  
Seq primer: mob.BGA+ET  
High quality sequence stop: 185.  
FEATURES  
Location/Qualifiers  
1..303  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:333219"  
/dev\_stage="19.5 dpc total fetus"  
/lab\_host="DH10B (ampicillin resistant)"  
/clone\_lib="Soares mouse p3NMFI9.5"  
/note="Vector: pT73D (Pharmacia) with a modified  
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand CDNA  
was primed with a Not I - oligo(dT) primer (5'  
TGTTCACATCTGAGTGGAGCGCGCATTTTGTGCTGCTCTCTGGAGACAGCTTCATATGCTCT 3',  
double-stranded CDNA was size selected, ligated to Eco RI  
adapters (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of a modified pT73 vector  
(Pharmacia). Library went through one round of  
normalization to a C<sub>0</sub>t = 5. Library constructed by Bento  
Soares and M.Fatima Bonaldo. RNA was kindly provided by  
Dr. Minoru Ko (Wayne State University)."

## ORIGIN

Query Match 6.9%; Score 46.6; DB 7; Length 303;  
Best Local Similarity 92.5%; Pred. No. 0.71;  
Matches 49; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ATGGGACCTGGATTTTGTGCTGCTCTCTGGAGACAGCTTCAGTATGCT 53  
DB 70 ATGGGACCTGGATTTTGTGCTGCTCTCTGGAGACAGCTTCATATGCT 122

RESULT 4  
LOCUS W40649 430 bp mRNA linear EST 11-SEP-1996  
DEFINITION mc43h12.r1 Soares mouse p3NMFI9.5 Mus musculus CDNA clone

ACCESSION	IMAGE:351335.5*	similar to gb:M10095 Mouse amelogenin (MOUSE); , MUS sequence.
VERSION	W40649	
KEYWORDS	W40649.1	GI:1324982
SOURCE	EST.	
ORGANISM	Mus musculus	(house mouse)
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 430)	
AUTHORS	Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubucq, T., Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, R., Steptoe, M., Tan, F., Underwood, K., Moore, B., Thaisang, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.	
TITLE	The WashU-HHMI Mouse EST Project	
JOURNAL	Unpublished (1996)	
COMMENT	Contact: Marra M/Mouse EST Project WashU-HHMI Mouse EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 850L, St. Louis, MO 63108 Tel.: 314 286 1800 Fax: 314 286 1810 Email: mouseest@wustl.wustl.edu This clone is available royalty-free through LNL; contact the IMAGE Consortium ( <a href="mailto:info@image.lnl.gov">info@image.lnl.gov</a> ) for further information. MG1:223135	
FEATURES	Seq primer: ETPprimer	
SOURCE	High quality sequence scrop: 323. Location/Qualifiers 1..430 /organism="Mus musculus" /mol_type="mRNA" /db_xref="taxon:10090" /clone="IMAGE:351335" /dev_stage="19.5 dpc total fetus" /lab_host="MD10B (ampicillin resistant)" /clone_lib="Soares mouse p3MMF19.5" /note="Vector: pVT73D (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5], TGTTACCACTGAGCGAGGGAGCGGCCCATTTTTTTTTTTT 3'), tetrac-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pVT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M.Patima Bonaldi. RNA was kindly provided by Dr. Minoru Ko (Wayne State University)."	
ORIGIN		
Query Match	6.9%; Score 46.6;	DB 7; Length 430;
Best Local Similarity	92.5%;	Pred. No. 0.75;
Matches	49; Conservative	0; Mismatches 4; Indels 0; Gaps 0;
Dy	1 ATGGGACCTGCATTTGTTCCTCCTCCTCGGAGACCCCTTCATATGCC 53 	58 ATGGGACCTGCATTTGTTCCTCCTCCTCGGAGACCCCTTCATATGCC 110 
LOCUS	W36345	536 bp mRNA linear EST 11-SEP-1996
DEFINITION	mbJ2c12.r1 Soares mouse p3MMF19.5 Mus musculus cDNA clone IMAGE:351335.5*, similar to gb:M10095 Mouse amelogenin (MOUSE); , mRNA sequence.	
ACCESSION	W36345	GI:1318120
VERSION	W36345.1	
KEYWORDS	EST.	
SOURCE	Mus musculus (house mouse)	
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 536)	Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Giesel, S., Kucaba, T., Lacy, M., Le, M., Martin, V., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.	The WashU-HMI Mouse EST Project	Unpublished (1996)	
	Contact: Marra M/Mouse EST Project			
	WashU-HMI Mouse EST Project			
	Washington University School of Medicine			
	4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108			
	Tel: 314 286 1800			
	Fax: 314 286 1810			
	Email: mouseest@watson.wustl.edu			
	This clone is available royalty-free through LINT ; contact the			
	IMAGE Consortium (info@image.llnl.gov) for further information.			
	MG1:216366			
	Seq primer: ETPPrimer			
	High quality sequence stop: 359.			
	Location/Qualifiers			
	1..536			
	/organism="Mus musculus"			
	/mol_type="mRNA"			
	/db_xref="taxon:10090"			
	/clone="IMAGE:334966"			
	/dev_stage="19.5 dpc total fetus"			
	/lab_host="DH10B (ampicillin resistant)"			
	/clone_1lb="Soares mouse p3NMF9.5"			
	/note="Vector: pT73D (Pharmacia) with a modified			
	polylinker; site_1: Not I; Site 2: Eco RI; 1st strand cDNA			
	was primed with a Not I - 0190(dt) primer [5,			
	TGTTACCAATCTGAAGTGGGAGCGGCGCATTTTCTTTTCTTTT 3']			
	double-stranded cDNA was size selected, ligated to Eco RI			
	adapters (Pharmacia), digested with Not I and cloned into			
	the Not I and Eco RI sites of a modified pT73 vector			
	(Pharmacia). Library went through one round of			
	normalization to a Cot = 5. Library constructed by Bento			
	Soares and M.Fatima Bonaldo. RNA was kindly provided by			
	Dr. Minoru Ko (Wayne State University)."			
ORIGIN				
Query Match	6.9%; Score 46.6; DB 7; Length 536;			
Best Local Similarity	92.5%; Pred. No. 0.78;			
Matches	49; Conservative 0; Mismatches 4; Indels 0; Gaps 0;			
OR				
1 ATGGGAGACCTGGATTTGTTGCTGCTGCTGGAGAGAGCCTCAGATGCC 53				
53 ATGGGAGACCTGGATTTGTTGCTGCTGCTGGAGAGAGCCTTGCATGCC 105				
RESULT 6				
AY419441	585 bp DNA linear GSS 17-DEC-2003			
LOCUS	AY419441			
DEFINITION	Mus musculus AMELX gene, VIRTUAL TRANSCRIPT, partial sequence,			
GENOMIC SURVEY SEQUENCE.				
ACCESSION	AY419441			
VERSION	AY419441.1 GI:39775398			
KEYWORDS	GSS.			
SOURCE	Mus musculus (house mouse)			
ORGANISM	Mus musculus			
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;				
1 (bases 1 to 585)				
Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejariwal, A.,				
Todd, M.A., Tanenbaum, D.M., Clavello, D.R., Lu, F., Murphy, B.,				
Ferriter, S., Wang, G., Zheng, X.H., White, T.J., Smitsky, J.C.,				
Adams, M.D. and Cargill, M.				
Interfering nonneutral evolution from human-chimp-mouse orthologous				
gene trios				
Science 302 (5652), 1960-1963 (2003)				
JOURNAL	14671302			
PUBMED				
REFERENCE	2 (bases 1 to 585)			

**AUTHORS** Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejaritwal, A., Todd, M.A., Tanenbaum, D.M., Civeille, D.R., Lu, F., Murphy, B., Ferriere, S., Wang, G., Zheng, X.H., White, T.J., Shtinsky, O.J., Adams, M.D., and Cargill, M.

**TITLE** Direct Submision

**JOURNAL** Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA

**COMMENT** These sequences were made by sequencing genomic exons and ordering them based on alignment.

**FEATURES** Location/Qualifiers

**source** 1..585

**gene** /organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
<1..585  
/gene="AMELX"  
/locus\_tag="HOM6891"

**ORIGIN**

Query Match 6.9%; Score 46.6; DB 9; Length 585;  
Best Local Similarity 92.5%; Pred. No. 0.79; Indels 0; Gaps 0;  
Matches 49; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

**QY** 1 ATGGGACCTGATTTGTTGCTGCTGCTGCGAGACGCTTACATATGCC 53  
1 ATGGGACCTGATTTGTTGCTGCTGCTGCGAGACGCTTATGCTATGCC 53

**Db** 1 ATGGGACCTGATTTGTTGCTGCTGCTGCGAGACGCTTATGCTATGCC 53

**RESULT 7** CD773419 645 bp mRNA linear EST 02-JUL-2003

**LOCUS** AGENCOURT 14713373 NIH MGC 190 Mus musculus cDNA clone

**DEFINITION** IMAGE:30501267 5', mRNA sequence.

**ACCESSION** CD773419

**VERSION** CD773419.1 GI:32431921

**KEYWORDS** EST.

**SOURCE** Mus musculus (house mouse)

**ORGANISM** Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

**REFERENCE** 1 (bases 1 to 645)  
NIH-MGC <http://mgs.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)

**AUTHORS** Contact: Daniela S. Gerhard, Ph.D.  
Office of Cancer Genomics  
National Cancer Institute / NIH  
Bldg. 31 Rm10A07 Bethesda, MD 20892  
Email: cgabs-r@mail.nih.gov

**TITLE** Tissue Procurement: Yoshitake Yamada, Takashi Nakamura, NIDCR

**JOURNAL** CDNA Library Preparation: CLONTECH Laboratories, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

**COMMENT** DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>  
Plate: NDCM23 row: k column: 04  
High quality sequence stop: 611.  
Location/Qualifiers

**FEATURES** 1..645

**source** /organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:30501267"  
/lab\_host="DH10B (phage-resistant)"  
/lab\_host="NIH\_MGC\_190"  
/note="Organ: Pooled - Molat; Vector: pDNR-LIB, Site 1: SfiI (ggccatcagcc); Site 2: SfiI (ggccgctcgcc); Non-normalized full-length enriched library 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGCCATTAAGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCCGAGCGCGCAGCATG-dt(30)BN-3' (where B = A, C, G and N = A, C, G, or T). Average insert size 1.71

kb (range 0.5-3.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA) Corp."

**ORIGIN**

Query Match 6.9%; Score 46.6; DB 6; Length 645;  
Best Local Similarity 92.5%; Pred. No. 0.81; Indels 0; Gaps 0;  
Matches 49; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

**QY** 1 ATGGGACCTGATTTGTTGCTGCTGCTGCGAGACGCTTACATATGCC 53  
73 ATGGGACCTGATTTGTTGCTGCTGCTGCGAGACGCTTATGCTATGCC 125

**Db** 73 ATGGGACCTGATTTGTTGCTGCTGCTGCGAGACGCTTATGCTATGCC 125

**RESULT 8** CB056709 707 bp mRNA linear EST 17-JAN-2003

**LOCUS** NISC J1905.w1 Soares NMBP13-15 Mus musculus cDNA clone

**DEFINITION** IMAGE:4848584 5', mRNA sequence.

**ACCESSION** CB056709

**VERSION** CB056709.1 GI:27794996

**KEYWORDS** EST.

**SOURCE** Mus musculus (house mouse)

**ORGANISM** Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

**REFERENCE** 1 (bases 1 to 707)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
Unpublished (1997)

**JOURNAL** Email: cgabs-r@mail.nih.gov

**COMMENT** CDNA Library Preparation:  
DNA Sequencing by: The I.M.A.G.E. Consortium/LNL  
Sequencing Center (NISC)  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>  
Plate: LLM10779 row: N column: 9  
Seq primer: T7 primer.  
Location/Qualifiers

**FEATURES** 1..707

**source** /organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:4848584"  
/issue\_type="pituitary gland"  
/dev\_stage="juvenile, 13-15 days"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="Soares NMBP13-15"  
/note="Organ: brain; Vector: pT73D-Paci, Site 1: NotI, Site 2: BcoRI; 1st strand cDNA was primed with a NotI-oligo(dT) primer  
5'-AACTGAGAGATTCGGCCGCTGATCCAGATGTTTTTTTTTTTTTT-3'; double-stranded cDNA was ligated to EcoRI adaptors 5'-AATCGGACGAGG-3' and 5'-CTGTGCGG-3' (Pharmacia), digested with NotI and cloned into the NotI and BcoRI sites of the pT73D-Paci vector. Library went through one round of normalization, and was constructed in the laboratory of M. Bento Soares (University of Iowa)."

**ORIGIN**

Query Match 6.9%; Score 46.6; DB 6; Length 707;  
Best Local Similarity 92.5%; Pred. No. 0.82; Indels 0; Gaps 0;  
Matches 49; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

**QY** 1 ATGGGACCTGATTTGTTGCTGCTGCTGCGAGACGCTTACATATGCC 53  
47 ATGGGACCTGATTTGTTGCTGCTGCTGCGAGACGCTTATGCTATGCC 99





ORIGIN

Query Match 6.9%; Score 46.6; DB 6; Length 843;  
Best Local Similarity 92.5%; Pred. No. 0.85;  
Matches 49; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

1 ATGGGACCTGATTTGTTGCTGCTGCTGAGAGACCTTGCATATGCC 53  
|||||  
40 ATGGGACCTGATTTGTTGCTGCTGCTGAGAGACCTTGCATATGCC 92  
|||||

Corp. Note: this is a NIH\_MGC Library."

/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:30293239"  
/issue\_type="embryonic limb, maxilla and mandible"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_136"  
/note="Vector: PCMV-SPORT6.1; Site.1: EcoRV; Site.2: NotI;  
Normalized, full-length enriched library from pool of  
mouse embryonic limb, maxilla and mandible, embryonic day  
17.5, 18.5 and newborn (mandible (5, 4 and 1 limb and jaw  
equivalents from respective days). Cloned directionally,  
oligo-dT primed (5'-GACTACTTTCATGTCGAGCGGCCGCCCTT)15-3'.  
Size selected for the >1kb fragments, average insert size  
1.2 kb. Normalization to Cot 7.5. Tissue contributed by  
David Rowe; library constructed by Resgen, Invitrogen  
Corp. Note: this is a NIH\_MGC Library."

RESULT 11  
CB588525 859 bp mRNA linear EST 03-APR-2003  
LOCUS IMAGE:30289751 5', mRNA sequence.  
DEFINITION CB588525  
ACCESSION CB588525.1 GI:29506381  
VERSION EST.  
KEYWORDS Mus musculus (house mouse)  
SOURCE Mus musculus  
ORGANISM Mus musculus (house mouse)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 859)  
NIH-MGC http://img.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-remail.nih.gov  
Tissue Procurement: Dr. David Rowe  
CDNA Library Preparation: Invitrogen Corp  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: NDAM329 row: e column: 24  
High quality sequence stop: 763.  
Location/Qualifiers  
1. 859  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:30289751"  
/issue\_type="embryonic limb, maxilla and mandible"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_136"  
/note="Vector: PCMV-SPORT6.1; Site.1: EcoRV; Site.2: NotI;  
Normalized, full-length enriched library from pool of  
mouse embryonic limb, maxilla and mandible, embryonic day  
17.5, 18.5 and newborn (mandible (5, 4 and 1 limb and jaw  
equivalents from respective days). Cloned directionally,  
oligo-dT primed (5'-GACTACTTTCATGTCGAGCGGCCGCCCTT)15-3'.  
Size selected for the >1kb fragments, average insert size  
1.2 kb. Normalization to Cot 7.5. Tissue contributed by  
David Rowe; library constructed by Resgen, Invitrogen

FEATURES  
SOURCE

ORIGIN

Corp. Note: this is a NIH\_MGC Library."

Query Match 6.9%; Score 46.6; DB 6; Length 859;  
Best Local Similarity 92.5%; Pred. No. 0.85;  
Matches 49; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

1 ATGGGACCTGATTTGTTGCTGCTGCTGAGAGACCTTGCATATGCC 53  
|||||  
59 ATGGGACCTGATTTGTTGCTGCTGCTGAGAGACCTTGCATATGCC 111  
|||||

RESULT 12  
CB589177 865 bp mRNA linear EST 03-APR-2003  
LOCUS IMAGE:30290534 5', mRNA sequence.  
DEFINITION CB589177  
ACCESSION CB589177.1 GI:29507033  
VERSION EST.  
KEYWORDS Mus musculus (house mouse)  
SOURCE Mus musculus  
ORGANISM Mus musculus (house mouse)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 865)  
NIH-MGC http://img.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-remail.nih.gov  
Tissue Procurement: Dr. David Rowe  
CDNA Library Preparation: Invitrogen Corp  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: NDAM331 row: f column: 15  
High quality sequence stop: 630.  
Location/Qualifiers  
1. 865  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:30290534"  
/issue\_type="embryonic limb, maxilla and mandible"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_136"  
/note="Vector: PCMV-SPORT6.1; Site.1: EcoRV; Site.2: NotI;  
Normalized, full-length enriched library from pool of  
mouse embryonic limb, maxilla and mandible, embryonic day  
17.5, 18.5 and newborn (mandible (5, 4 and 1 limb and jaw  
equivalents from respective days). Cloned directionally,  
oligo-dT primed (5'-GACTACTTTCATGTCGAGCGGCCGCCCTT)15-3'.  
Size selected for the >1kb fragments, average insert size  
1.2 kb. Normalization to Cot 7.5. Tissue contributed by  
David Rowe; library constructed by Resgen, Invitrogen  
Corp. Note: this is a NIH\_MGC Library."

FEATURES  
SOURCE

ORIGIN

Query Match 6.9%; Score 46.6; DB 6; Length 865;  
Best Local Similarity 92.5%; Pred. No. 0.85;  
Matches 49; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

1 ATGGGACCTGATTTGTTGCTGCTGCTGAGAGACCTTGCATATGCC 53  
|||||  
58 ATGGGACCTGATTTGTTGCTGCTGCTGAGAGACCTTGCATATGCC 110  
|||||

RESULT 13  
CB587051 886 bp mRNA linear EST 03-APR-2003  
LOCUS IMAGE:30280704 NIH\_MGC\_136 Mus musculus cDNA clone  
DEFINITION

IMAGE:30295123 5', mRNA sequence.  
CB587051  
CB587051.1 GI:29504907  
EST.  
SOURCE  
Mus musculus (house mouse)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 886)  
NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
Tissue Procurement: Dr. David Rowe  
CDNA Library Preparation: Invitrogen Corp  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>  
Plate: NDAM343 row: e column: 20  
High quality sequence stop: 619.  
Location/Qualifiers  
1. .886  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:30295123"  
/tissue\_type="embryonic limb, maxilla and mandible"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_1lb="NIH\_MGC\_136"  
/note="Vector: PCMV-SPORE.1; Site.1: EcoRV; Site.2: NotI; Normalized, full-length enriched library from pool of mouse embryonic limb, maxilla and mandible, embryonic day 17.5, 18.5 and newborn (mandible (5, 4 and 1 limb and jaw equivalents from respective days). Cloned directionally, oligo-dT primed (5'-GACTACTCTAGATCGGAGCGGCCGCTT)15-3'. Size selected for the >1kb fragments, average insert size 1.2 kb. Normalization to Cot 7.5. Tissue contributed by David Rowe; library constructed by Resgen, Invitrogen Corp. Note: this is a NIH\_MGC Library."

ORIGIN

Query Match 6.9%; Score 46.6; DB 6; Length 886;  
Best Local Similarity 92.5%; Pred. No. 0.85;  
Matches 49; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ATGGGAGCCTGATTTGTTGCTGCTCTCGGAGACCTTCACTATGCC 53  
|||||  
84 ATGGGAGCCTGATTTGTTGCTGCTCTCGGAGACCTTTCCTATGCC 136  
|||||

RESULT 14  
CB589251 889 bp mRNA linear EST 03-APR-2003  
LOCUS  
DEFINITION  
AGENCOURT\_12780700 NIH\_MGC\_136 Mus musculus CDNA clone  
IMAGE:30287069 5', mRNA sequence.  
CB589251  
CB589251.1 GI:29507107  
EST.  
SOURCE  
Mus musculus (house mouse)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 889)  
NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
Tissue Procurement: Dr. David Rowe  
CDNA Library Preparation: Invitrogen Corp

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

FEATURES  
source  
1. .891  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:30294719"  
/tissue\_type="embryonic limb, maxilla and mandible"

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>  
Plate: NDAM322 row: f column: 06  
High quality sequence start: 23  
High quality sequence stop: 581.  
Location/Qualifiers  
1. .889  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:30287069"  
/tissue\_type="embryonic limb, maxilla and mandible"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_1lb="NIH\_MGC\_136"  
/note="Vector: PCMV-SPORE.1; Site.1: EcoRV; Site.2: NotI; Normalized, full-length enriched library from pool of mouse embryonic limb, maxilla and mandible, embryonic day 17.5, 18.5 and newborn (mandible (5, 4 and 1 limb and jaw equivalents from respective days). Cloned directionally, oligo-dT primed (5'-GACTACTCTAGATCGGAGCGGCCGCTT)15-3'. Size selected for the >1kb fragments, average insert size 1.2 kb. Normalization to Cot 7.5. Tissue contributed by David Rowe; library constructed by Resgen, Invitrogen Corp. Note: this is a NIH\_MGC Library."

ORIGIN

Query Match 6.9%; Score 46.6; DB 6; Length 889;  
Best Local Similarity 92.5%; Pred. No. 0.86;  
Matches 49; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ATGGGAGCCTGATTTGTTGCTGCTCTCGGAGACCTTCACTATGCC 53  
|||||  
78 ATGGGAGCCTGATTTGTTGCTGCTCTCGGAGACCTTTCCTATGCC 130  
|||||

RESULT 15  
CB574837 891 bp mRNA linear EST 02-APR-2003  
LOCUS  
DEFINITION  
AGENCOURT\_12972279 NIH\_MGC\_136 Mus musculus CDNA clone  
IMAGE:30294719 5', mRNA sequence.  
CB574837  
CB574837.1 GI:29494367  
EST.  
SOURCE  
Mus musculus (house mouse)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 891)  
NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
Tissue Procurement: Dr. David Rowe  
CDNA Library Preparation: Invitrogen Corp  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>  
Plate: NDAM342 row: d column: 24  
High quality sequence start: 19  
High quality sequence stop: 658.  
Location/Qualifiers  
1. .891  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:30294719"  
/tissue\_type="embryonic limb, maxilla and mandible"

/lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH\_MGC\_136"  
 /note="Vector: PCMV-SPORT6.1; Site 1: EcoRV; Site 2: NotI;  
 Normalized, full-length enriched library from pool of  
 mouse embryonic limb, maxilla and mandible, embryonic day  
 17.5, 18.5 and newborn (mandible (5' 4 and 1 limb and jaw  
 equivalents from respective days). Cloned directionally,  
 oligo-dT primed (5'-GACTAGTCTAGATCGGAGCGGCCGCC(7)15-3'.  
 Size selected for the >1kb fragments, average insert size  
 1.2 kb. Normalization to Cot 7.5. Tissue contributed by  
 David Rowe; library constructed by ResGen, Invitrogen  
 Corp. Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 6.9%; Score 46.6; DB 6; Length 891;  
 Best Local Similarity 92.5%; Pred. No. 0.86;  
 Matches 49; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ATGGGGAAGCTGATTTTGTTCCTGCTCTGGAGACAGCTTCAGTATGCC 53  
 |||||||  
 DB 79 ATGGGGAAGCTGATTTTGTTCCTGCTCTGGAGACAGCTTCAGTATGCC 131

Search completed: February 8, 2005, 22:18:22  
 Job time: 8100.44 secs